

HELSINGFORS UNIVERSITET  
HUMANISTISKA FAKULTETEN

## **Bedömningen av muntliga språkkunskaper i svenska**

En analys av bedömarnas reflektioner kring elevbedömningens reliabilitet

Finska, finskugriska och nordiska institutionen

Pro gradu-avhandling i nordiska språk

Sivi Lindroos

Handledare: Hanna Lehti-Eklund

17.3. 2010

## Innehåll

1 Inledning .....	1
1.1 Projektet HY-TALK vid Helsingfors universitet .....	2
1.2 Syfte .....	4
1.3 Avhandlingens disposition .....	5
2 Material och metod .....	6
2.1 Bedömningssession 1 .....	6
2.2 Bedömningssession 2 .....	7
2.3 Metod .....	7
3 Styrdokument .....	9
3.1 Europeisk referensram för språk .....	9
3.2 Användning av CEFR .....	10
3.3 Utmaningar i CEFR-skalornas användning .....	12
3.4 Den finländska nivåskalan för språkkunskaper .....	14
3.5 KORU-projektet inom de finska yrkeshögskolorna samt språkcenter vid finska universitet .....	16
4 Forskningsöversikt .....	18
4.1 Bedömaren i fokus .....	18
4.2 Muntliga språkprov .....	19
4.3 Bedömare .....	22
4.4 Bedömning av muntlig språkfärdighet .....	23
4.5 Teoretiska modeller för språkbedömning .....	24
4.6 Principer i språkbedömningen .....	30
4.6.1 Bedömningens validitet .....	31
4.6.2 Bedömningens reliabilitet .....	32
4.6.3 Interbedömarreliabilitet och intrabedömarreliabilitet .....	33
4.6.4 Autenticitet .....	34
4.7 Andra faktorer i språkbedömningen .....	36
4.8 Den europeiska referensramens syn på bedömning .....	37
4.9 Språkkunskapernas delkomponenter i en analytisk bedömning .....	39
5 Resultat .....	42
5.1 Bedömare i fokus .....	42
5.2 De viktigaste resultaten av intervjuerna .....	60
5.2.1 Reliabilitetsfrågan .....	60
5.2.2 Kommunikation och interaktion i provsituationen .....	61
5.2.3 Grammatikens roll i muntliga språkkunskaper .....	62
5.3 Reliabilitet av bedömning enligt nivåskalan för språkfärdighet .....	63
6 Diskussion av resultaten .....	67
6.1 Hur borde man gå vidare? .....	69
7 Avslutning .....	72
Litteratur .....	74
 Bilaga 1: Intervjufrågor .....	77
Bilaga 2: Skalan för muntlig färdighet i läroplansgrunderna .....	78
Bilaga 3: HY-TALK Hösten 2007, gymnasienivå .....	82
Bilaga 4: HY-TALK Högskolenivå .....	84

## Tabeller

Tabell 1: Bedömningssessionerna.....	6
Tabell 2: CEFR-skalan och nivåskalan för språkkunskaper. ....	14
Tabell 3: Färdighetsnivåerna i svenska som andra inhemska språket på gymnasienivå. ...	15
Tabell 4: Färdighetsnivån på gymnasienivå, kriterierna för Tal på nivå B1.2.....	15
Tabell 5: Jämförelse av kravnivån i svenska vid högskolor i jämförelse med andra beskrivningar av språkkunskaper (Elsinen & Juurakko-Paavola 2006: 92). Min översättning.....	17
Tabell 6: Bedömningssätt.....	38
Tabell 7: Fonologisk kontroll, beskrivning av fonologisk kompetens enligt Europeisk referensram för språk (2007: 113) .....	40
Tabell 8: Muntlig flyt enligt Europeisk referensram för språk (2007) .....	41
Tabell 9: Bedömarprofil, bedömare 1. ....	43
Tabell 10: Bedömarprofil, bedömare 2 .....	46
Tabell 11: Bedömarprofil, bedömare 3 .....	48
Tabell 12: Bedömarprofil, bedömare 4 .....	51
Tabell 13: Bedömarprofil, bedömare 5 .....	53
Tabell 14: Bedömarprofil, bedömare 6 .....	55
Tabell 15: Bedömarprofil, bedömare 7 .....	57
Tabell 16: Helhetsbedömning, gymnasienivå.....	63
Tabell 17: Nivåskalorna i siffror .....	64
Tabell 18: Statistiskt test av resultaten på gymnasienivå.....	65
Tabell 19: Helhetsbedömning, högskolenivå.....	65
Tabell 20: Statistiskt test av resultaten på högskolenivå.....	67

## Figurer

Figur 1: Performance-based assessment (McNamara 1996: 9). Min översättning. ....	19
Figur 2: Modell över den kommunikativa språkförmågan (efter Eklund Heinonen 2009: 48) .....	29
Figur 3: Autenticitet (ibid. 1996: 23) .....	36

# 1 Inledning

Det finns en hel del utmaningar vid bedömning av muntlig språkfärdighet. Bedömarens subjektiva betoningar och elevernas personligheter blir lätt synliga i bedömningen. Svensklärarna i Finland har upplevt bedömningen av muntliga språkkunskaper subjektivare än bedömningen av skriftliga språkkunskaper (Elsinen & Juurakko-Paavola 2006: 63). Under de senaste åren har man i Europa utvecklat nya sätt att kriteriebaserad bedöma muntliga språkkunskaper. Den här avhandlingen handlar om sju bedömares subjektiva syn på kriteriebaserad bedömning av språkkunskaper i svenska. Arbetets syfte är att analysera bedömarens reflektioner av att bedöma muntliga kunskaper i svenska. Betoningarna i bedömningen jämförs med bedömningsresultaten i pilotundersökningen i projektet HY-TALK. Varje bedömare har sitt eget sätt att bedöma, men trots detta tycks de bedömningskriterier som har använts fungera relativt bra i elevbedömningen.

*The Common European Framework of Reference for Languages: Learning, teaching, assessment* (CEFR), publicerad år 2001, fungerar som bas till den nya läroplanen för språkundervisningen i finska grundskolor och gymnasier. I Finland har man utvecklat en nivåskala för språkfärdigheter på basis av CEFR, och denna skala används i evalueringen av språkfärdigheter i grund- och gymnasieutbildningen. Europeisk referensram för språk (2007), eller på finska *Eurooppalainen viitekehys: Kielten oppimisen, opettamisen ja arvioinnin yhteinen eurooppalainen viitekehys* (2003) är från en finländsk synvinkel ett relativt nytt, fastän långt utarbetat, europeiskt sätt att planera, utveckla och evaluera språkundervisning, ställa krav på den och ställa kriterier på elevernas språkkunskaper i olika faser i språkinläringen. Referensramen är ett kriteriebaserat verktyg för språkevalueringen. Avsikten med den är att standardisera och förenhetliga evalueringen av språkkunskaper i de europeiska länderna.

De nya läroplansgrunderna och CEFR betonar evalueringen av funktionella språkkunskaper. Då beaktas också muntlig färdighet på ett nytt sätt i jämförelse med

tidigare läroplansgrunder. I det finska språkundervisningspolitiska projektets (KIEPO) slutrapport konstateras, att evalueringen av funktionella språkfärdigheter medför också ett behov för att evaluera muntliga färdigheter. Det skulle vara viktigt att ta muntliga språkfärdigheter med i evalueringen till exempel i studentskrivningarna för att betoningarna i läroplansgrunderna skulle vara mera synliga i praktiken i undervisningen. (Tarnanen, Huhta & Pohjala 2007: 404)

Inlärarens muntliga kompetens kräver uppmärksamhet också enligt Green-Vänttinen och Rosenberg-Wolff (2007: 83-85). Den nyaste läroplanen för gymnasiet (2003) betonar muntlig kommunikation i undervisningen, och eftersom läroplanen fungerar som bas för undervisningen, är det viktigt att evalueringen av muntliga språkfärdigheter betonas också i praktiken.

Muntliga språkfärdigheter i skolundervisningen är ett aktuellt tema, och därför har jag valt att forska i det. Enligt min åsikt är muntlig kompetens en ytterst viktig del av språkfärdigheter och den behöver därför uppmärksamhet. Evalueringen av muntliga språkfärdigheter behöver ännu utvecklas och standardiseras. Med hjälp av evalueringen kan man också utveckla själva undervisningen. Lärare, bedömare och språkstuderande har ett verktyg för evaluering av muntliga språkfärdigheter: den europeiska referensramen för språk. Det är ett nytt verktyg i Finland och det behövs övning för att kunna använda och tillämpa den (Kalliomäki 2007: 42).

## **1.1 Projektet HY-TALK vid Helsingfors universitet**

Projektet HY-TALK är ett tvärvetenskapligt projekt vid Helsingfors universitet och görs i samarbete mellan Institutionen för lärarutbildning, Institutionen för moderna språk samt Finska, finskugriska och nordiska institutionen. Projektet syftar till att förstärka utbildningen och bedömningen av muntliga språkkunskaper i främmande språk i den allmänna utbildningen och på högskolenivå.

Den teoretiska forskningen gäller frågor kring evaluering av muntliga språkfärdigheter. Projektets syfte är att utveckla undervisning, studier och inläring av muntliga

språkfärdigheter på de olika utbildningsnivåerna genom att validera bedömningsresultatens pålitlighet. Man vill nå syftet genom att teoretiskt och empiriskt validera nivåskalan för språkfärdighet.

Den muntliga färdigheten bedöms i dialogiska prov. Material samlas in från årskurs sju i grundskolan, från årskurs ett i gymnasiet och från elever på högskolenivå. Bedömningen görs både analytiskt och holistiskt, och vid den analytiska bedömningen beaktas fyra kriterier: uttal, flyt, ordförråd (bredd) och grammatisk korrekthet. Praktiskt sett har projektet för svenska språkets del betytt inspelningar av elevprestationer på olika nivåer. Inspelningar på gymnasienivå gjordes hösten 2007 och på högskolenivå hösten 2008. Efter inspelningarna har det ordnats bedömningssessioner för bedömare då de sett elevernas prestationer på band och evaluerat dem enligt nivåskalan för språkfärdighet. Jag har varit med i projektet som forskningsassistent och deltagit i inspelningarna samt i bedömningarna av elevprestationer. De flesta bedömare är också mina informanter som jag individuellt har intervjuat efter bedömningssessionerna.

När det gäller evalueringen av tyska och engelska inspelningar, har man redan fått vissa resultat av HY-TALK-bedömningarna. Bedömningsresultaten i tyska hade en reliabilitetsnivå på 0,88 mellan bedömarna, vilket betyder att interbedömarreliabiliteten kvantitativt sett var ganska hög utgående från bedömningsresultaten. De största skillnaderna mellan bedömarna kom fram i bedömningen av uttal och ordförråd. Det fanns inga signifikanta skillnader mellan bedömare när det gäller genomförande av uppgiften och flyt enligt en ensidig variansanalys, där man undersökte om skillnaderna i bedömningsresultaten är större mellan än inom bedömarna (bevis på dålig interbedömarreliabilitet). En kvalitativ analys av bedömarens kommentar avslöjade att många bedömare evaluerade att elevernas uttal var på en högre nivå på skalan än de andra delkomponenterna i språkkunskaperna. Forskarna konstaterar att det kan finnas skäl till att förändra beskrivningarna för uttalet i nivåskalans kriterier. Bedömare problematiserade också flytets definition. Det kom också fram att bedömare jämförde elever med varandra i bedömningen och tänkte att till exempel parets prestation påverkade elevens performans i provsituationen. (Hildén & Martikainen 2009a) För

engelskans del ser det ut som variansen vore större mellan bedömarna och att reliabiliteten därmed vore lägre (Hildén & Martikainen 2009b).

## 1.2 Syfte

Syftet med mitt pro gradu-arbete är att undersöka hur två expertgrupper bestående av lärare bedömer elevernas muntliga kompetens i svenska som andraspråk på gymnasienivå samt på högskolenivå. Den ena gruppen som består av fyra universitetslärare bedömer gymnasieelevernas muntliga färdigheter i A-svenska enligt nivåskalan för språkfärdigheter i läroplansgrunderna, som baserar sig på den europeiska referensramen för språk. Den andra gruppen som består av tre universitetslärare bedömer högskoleelevernas muntliga färdigheter i svenska. Min undersökning syftar på att redogöra hur bedömarna upplever bedömningen, och vilka saker de tycker är viktiga i bedömningen av muntliga språkfärdigheter. Jag vill veta, vad de avser om användning och pålitlighet av den europeiska referensramen för språk samt frågor relaterade till dess reliabilitet och validitet. Jag är också intresserad av bedömarnas egna åsikter om och attityder till användningen av referensramen i evaluering.

Jag kopplar ihop bedömarnas åsikter om bedömningen med resultaten av två praktiska bedömningssessioner. I den första bedömer fyra informanter sex elevpar på gymnasienivå enligt nivåskalan i läroplansgrunderna. I den senare bedömer tre informanter fem elevpar på högskolenivå enligt samma nivå, som alltså används i HY-TALK också på högskolenivån. Jag vill ta reda på om bedömningen har bra reliabilitet, dvs. är enhetlig i praktiken, eller om elevernas resultat varierar mellan bedömarna. Jag ska undersöka om bedömarnas subjektiva betoningar i bedömningen är synliga i bedömningsresultaten så att en bedömare som tar upp synpunkter på sin bedömning under intervjun bedömer i olika än andra bedömare med olika åsikter

Mina forskningsfrågor är följande:

1. Hur upplever två grupper av bedömare bedömningen av elevernas muntliga språkkunskaper i svenska enligt nivåskalan för språkfärdighet?

## 2. Hurdan är reliabiliteten mellan bedömnarna i bedömningen enligt nivåskalan?

Min forskning är alltså tudelad: för det första är jag intresserad av bedömnarnas betoningar och åsikter om problem som kan uppstå vid bedömningen av muntliga språkfärdigheter enligt den europeiska referensramen. För det andra vill jag undersöka om det i praktiken finns problem med reliabiliteten i den bedömning som mina informanter utför.

Jag har inga fasta hypoteser, men jag antar att det finns några, större eller mindre problem i bedömningen. Som bas för detta antagande använder jag Kalliomäkis (2007) pro gradu-avhandling. Kalliomäki skriver i sin pro gradu-avhandling att det uppstår problem när det gäller elevbedömning enligt den europeiska referensramen (2003). Enligt hennes undersökning upplever lärarna skalorna som för breda och därför svåra att använda (ibid. 28, 35). Tre av fem lärare som hon intervjuade använde inte skalorna i sitt arbete. Jag antar att det uppstår också några slags problem i mitt material, fastän mina informanter inte jobbar i gymnasier eller i grundskolor utan vid en högskola. Jag ska ta reda på hurdana åsikter informanterna har om användningen av referensramen och hur de upplever evalueringen av muntliga språkkunskaper enligt skalan. Fokus är på hur skalan och dess tillämpning fungerar i praktiken. Kalliomäkis (2007) informanter är vanliga språklärare i grundskolor och gymnasier, och det kom fram att bara en del av lärarna hade använt CEFR-baserade skalor i evalueringen av muntliga språkfärdigheter. Jag skall ta med sådana lärare som är bekanta med skalorna och har använt dem i sitt arbete som lärare. Mina informanter är experter i evaluering av språkfärdigheter och kan också använda CEFR-baserade skalor.

### **1.3 Avhandlingens disposition**

Avhandlingen är disponerad på följande sätt. Ovan har jag berättat om uppsatsens syfte samt forskningsfrågorna. I följande kapitel redogör jag för materialet och metoderna för min undersökning. Kapitel 3 beskriver undersökningens styrdokument, CEFR och den finska nivåskalan för språkfärdigheter. I kapitel 4 skall jag presentera forskning kring språkbedömning och fokusera speciellt på bedömningen av muntliga språkfärdigheter.



Viktiga begrepp i språkbedömningen blir autenticitet, validitet, reliabilitet samt kommunikativ kompetens. I femte kapitlet presenterar jag mina resultat. Kapitel sex innehåller diskussion av resultaten. I det sista kapitlet blickar jag tillbaka samt ger förslag för vidare forskning.

## 2 Material och metod

Mitt material består av intervjuer av sju universitetslärare. Som sekundärmaterial använder jag också elevbedömningar från två bedömningssessioner inom projektet HY-TALK vid Helsingfors universitet (se tabell 1). Alla lärarna har deltagit i en bedömningssession och bedömt de muntliga prestationerna i dem, varefter jag har intervjuat dem. Alla bedömare, dvs. universitetslärarna först bedömer elever i HY-TALK enligt nivåskalan för språkfärdighet i läroplansgrunderna.

**Tabell 1: Bedömningssessionerna**

Bedömningssession	Deltagare	Material	Tidpunkt
Bedömningssession 1: gymnasienivå	Bedömare 1, 2, 3 och 4	12 elevprestationer på gymnasienivå, första årskurs	Elevprestationer bandades in på hösten 2007, bedömningssessionen på våren 2008
Bedömningssession 2: högskolenivå	Bedömare 1, 4 samt bedömare 5, 6 och 7	10 elevprestationer på högskolenivå	Studentprestationer bandades in på hösten 2008, bedömningssession på våren 2009

### 2.1 Bedömningssession 1

Fyra lärare har tagit del i en pilotbedömningssession, där de har visats en videoinspelning av sex elevpar på gymnasienivå. Gymnasieeleverna har utfört muntliga paruppgifter som har planerats inom projektet HY-TALK på Helsingfors universitet. Bedömningsskalan finns som bilaga (se bilaga 2). Lärarna har bedömt på basis av videoinspelningen. Inspelningen har visats till lärarna en gång. I början av bedömningssessionen har det berättats att i slutskedet av årskurs nio borde eleven vara på nivå A2.1 i muntlig förmåga för att få betyg 8. Lärarna har använt en

bedömningsblankett, där det finns olika aspekter i bedömningen som de måste bedöma separat. De har bedömt elevens genomförande av uppgiften, flyt, uttal, bredd, korrekthet och interaktion med skilda poäng. Till sist ger de en helhetsbedömning av elevens utförande. Eleverna har utfört parvis fyra uppgifter som är i olika grad styrda och innehåller kommunikativa aspekter (se bilaga 3). I den första uppgiften skall elever individuellt presentera sig själva på svenska. I den andra uppgiften diskuterar elever parvis i två vardagliga situationer. Replikerna är färdigt givna på finska. Uppgift tre är friare och innehåller planering av en utflykt på svenska med paret. De innehållsmässiga punkter som eleverna skall uttrycka står ändå på en lista. I uppgift fyra skall eleverna diskutera sina hobbyer med paret. Inga färdiga repliker ges.

## **2.2 Bedömningssession 2**

Fem av mina informanter har deltagit i en likadan bedömningssession av fem studentpar på högskolenivå på våren 2009. Lärarna har använt en likadan bedömningsblankett och bedömt samma 7 element i studentprestationerna. Två av de informanter (1 och 4) som har deltagit i bedömningssessionen på högskolenivå är desamma som vid bedömningssession 1. De studenter som har bedömts har parvis utfört fyra uppgifter som är också styrda och innehåller kommunikativa aspekter som till exempel samtal med en infödd talare (se bilaga 4). I uppgift 1 skall studenter presentera sig själva till en klasskamrat på utbytesorten. De saker som man skall nämna finns listade i uppgiftspapper. I uppgift 2 talar studenten med en infödd talare (inspelare) om ett ämne som man kan välja från listan. I uppgift tre skall man parvis diskutera en text som paret har läst på förhand. I den sista uppgiften skall man diskutera och motivera sina synpunkter gällande ett påstående som man väljer från en given lista. Uppgifterna är stort sett friare än motsvarande på gymnasienivå.

## **2.3 Metod**

Jag ska undersöka hur professionella lärare och bedömare bedömer elevernas muntliga färdigheter i ett andraspråk. Att bedöma språkkunskaper är en mångsidig och komplicerad process, och för att kunna fånga in på bästa möjliga sätt vad som styr

bedömarna när de bedömer och vilka tankar bedömningen väcker hos bedömarna, har jag valt kvalitativ intervju som min metod (Kvale 1997).

Efter intervjuerna har jag transkriberat samtalen för att kunna referera till materialet ordentligt. Lingvistisk transkribering behövs ändå inte eftersom det inte är väsentligt för mitt arbete. När man tematiserar och planerar en intervju måste man enligt Kvale (1997) fundera på tre grundläggande frågor: vad, varför och hur man intervjuar. Syftet med intervjuerna är att få svar på mina forskningsfrågor. Jag har inga fasta hypoteser, och har velat hålla intervjuerna öppna.

En kvalitativ intervju är öppen och flexibel (Kvale 1997: 82). Intervjuaren måste hålla reda på vad som är relevant och viktigt och försöka leda diskussionen i rätt riktning. Det finns ingen standardteknik eller regler för hur man gör en intervjuundersökning, så intervjuaren har själv allt ansvar för hur intervjun lyckas. Intervjusituationen är ostandardiserad, men hela undersökningen följer ändå oftast sju stadier: tematisering, planering, intervju, utskrift, analys, verifiering och rapportering (ibid. 85). Jag har börjat med att planera intervjuerna, sedan har jag intervjuat och transkriberat samtalen. Till sist följer analysen. Jag har analyserat intervjumaterialet med hjälp av meningskoncentreringsmetod (Kvale 1997: 175), så att jag har formulerat de meningar som intervjupersonerna uttryckt mera koncist med utgångspunkt i transkriberingarna. Långa uttalanden har jag omformulerat i några få ord för att sammanfatta och reducera intervjutexterna till kortare enheter.

Intervjuerna kan vara explorativa ”eller användas för prövning av hypoteser” (ibid. 94). En explorativ intervju är öppen och har lite struktur. Intervjuaren introducerar då ett område eller ett komplicerat problem som den som intervjuas försöker lösa. Intervjuaren följer upp personens svar och söker ny information om ämnet och nya synvinklar till området i fråga. Min öppna intervjumetod är lik detta. Jag har ingen färdig schablon eller bestämda svarsalternativ på de frågor jag ska ställa. Jag ställer frågor och tilläggsfrågor och frågar detaljer med utgångspunkt i de svar som jag får i intervjusituationen. Jag har ändå en färdigt planerat frågelista (se bilaga 1), och jag önskar få svar åtminstone på alla

de frågor som finns på listan. Eftersom jag intervjuar sju personer och är intresserad av just dessa personers svar, är det frågan om en fallstudie. Kvale (1997) nämner också postexperimentella studier, vilka innebär att intervjupersonerna utfrågas om hur de såg på experimentets utformning. I min undersökning är det fråga om en postexperimentell fallstudie: jag vill veta sju personers tankar om evalueringen av muntliga språkfärdigheter i svenska enligt skalan i läroplansgrunderna.

Jag är också intresserad av om det finns en korrelation mellan lärarnas åsikter och betoningar i bedömningen och deras bedömningsresultat. Jag presenterar bedömningsresultat med speciell fokus på bedömningsresultatens reliabilitet. Min undersökning innehåller således också en liten kvantitativ analys, men eftersom informanterna är sju stycken, måste den kvantitativa delen ses ha en stödande funktion till den kvalitativa analysen av reliabiliteten och intervjumaterialet som inte har någon statistisk signifikans.

### **3 Styrdokument**

#### **3.1 Europeisk referensram för språk**

Europarådet eller mera exakt rådet för kultursamarbete (Council for Cultural Co-operation, CDCC) som grundades år 1962 har utvecklat den europeiska referensramen som publicerades för första gången 1996 (Huttunen & Takala 2004: 326). Den har utarbetats av John Trim, Brian North, Daniel Coste och Joseph Sheils (Anderson (red.) 2002: 1). Andra versionen publicerades år 1998 och den senaste år 2001. Brian North har formulerat skalorna i referensramen (ibid. 7).

Europarådet har utvecklat språkutbildningen i Europa i över femtio års tid i form av olika projekt och haft som mål en internationell och nationell koherens i språkutbildningen. Synen på språkutbildningen är bred, och därför är det tillåtet och i några fall även viktigt att man tillämpar och tolkar Europarådets verktyg enligt nationella eller lokala behov. (Huttunen & Takala 2004: 325)

Före CEFR (2001) har man inom CDCC utvecklat andra skalor från och med 1970-talet. Till exempel "the Threshold Level" av van Ek och Trim (1990) har en kommunikativ inställning till språkundervisning och har påverkat kraftigt praktiken i skolor samt provplaneringen. CEFR baserar sig bland annat på Threshold Level och annan forskning samt andra arbeten inom CDCC. (CEFR 2001: 1). Kommunikativ kompetens har varit ett forskningsobjekt, och till exempel Bachmans kommunikativa språkförmåga, *communicative language ability*, CLA, anses vara en slags teoretisk bas för den europeiska referensramen.

### 3.2 Användning av CEFR

Starr Keddle (2004: 43) skriver att CEFR-skalan har både positiva och negativa sidor när det gäller dess användning på högstadie- och gymnasienivå (*secondary school*). Den fokuserar på situationell och funktionell språkförmåga samt på de strategier studenterna behöver när de använder språket: när de talar, skriver, lyssnar eller läser. Mekanisk inläring av grammatik betonas inte, utan inlärarspråket har mera betydelse. Detta leder till en mera kommunikativ språkanvändning. CEFR-skalan tar hänsyn till olika kommunikativa strategier och ger således läraren och inlärarna möjlighet att utveckla kompetens i olika strategier som till exempel turtagning, vilket lätt kan glömmas i klassrummet. Enligt Starr Keddle betonas referensramen också muntliga språkfärdigheter som vanligen ges mindre fokus i skolundervisningen. Genom ökad övning i muntliga färdigheter blir inläraren en mycket effektivare användare av målspråket, vilket tyder på att CEFR har en positiv washback-effekt, dvs. betoningen i en ny CEFR-baserad läroplan påverkar också de praktiska betoningarna i klassrummet.

Den europeiska referensramen för språk (Eurooppalainen viitekehys 2003: 8) har utvecklats för att användas som en gemensam europeisk bas för sammanställningen av studieplaner, läroplansgrunder, examina, läroböcker med mera. Den betonas språkförmåga i relation till kommunikation: den beskriver omfattande vad inläraren ska kunna för att kunna använda språket i kommunikation. Förhållningssättet till språkinläringen är funktionellt. Beskrivningen tar också hänsyn till språkets kulturella kontext. Referensramen definierar de nivåer, med vilka man kan mäta inlärarens

utveckling i språket. Referensramen skapar en gemensam bas för examina och kurser i Europa och befrämjar på så sätt deras genomskinlighet samt medverkar i internationellt samarbete och rörlighet mellan de europeiska länderna. För att kunna mäta och specificera språkkunskaper måste man dela upp språkförmågan i olika delfaktorer. Kunskaperna fungerar ändå i praktiken i en komplicerad interaktion. I referensramens första kapitel sägs det, att det är lärarnas och inlärarnas uppgift att förena dessa delfaktorer återigen till en dynamisk helhet (ibid. 19-20). Indelningen möjliggör också beskrivning av kommunikationsförmåga så att man kan ta hänsyn till endast ett område, till exempel läsförståelse.

Man har tagit CEFR i bruk till exempel i Katalonien och i Polen samt i Finland, till exempel i ett projekt vid Jyväskyläs universitet, kallat DIALANG (Anderson (red.) 2002: 7). Inom projektet har man använt skalorna i referensramen som bas i en undersökning där finska S2-lärare bords ordna delar i skalorna i referensramen i rätt ordning (Kaftandjieva & Takala 2002: 106). DIALANG strävar efter att utveckla ett Internet-baserat språkevalueringsystem som kan användas för självevaluering och självtestning. Skalorna för språkevalueringen inom DIALANG "will be expressed in terms of the Council of Europe (CoE) six-level scales of language proficiency" (Kaftandjieva & Takala 2002: 106). I undersökningen klarade lärarna av att ordna delarna i skalorna i en ordning som var lik originalen i CEFR, vilket innebär alltså att validiteten för att använda CEFR-skalor för att mäta språkfärdigheter är bra:

*"Summarizing the results of all applied methods of analysis (exploratory factor analysis, pattern matching, discriminant analysis and analysis of the agreement), we can conclude there is strong evidence in support of the construct validity of CEF scales of language proficiency." (Kaftandjieva & Takala 2002: 119).*

I Finland har också några högskolor tagit skalorna i den europeiska referensramen i bruk för att evaluera språkfärdigheter. Till exempel i ämnet nordiska språk vid Helsingfors universitet har man valt att använda dem som instrument i evalueringen av studenternas

muntliga färdighet (Green-Vänttinen & Rosenberg-Wolff 2007: 86). Där har man anpassat skalorna till ämnets ändamål.

När CEFR-skalorna skapades validerades de både kvalitativt och kvantitativt genom bedömning av elevprestationer på engelska, tyska och franska. Den kvalitativa valideringen gjordes i form av 32 sessioner med lärare som diskuterade bedömningen av elevparens muntliga språkfärdigheter. Meningen var för det första att få en uppfattning om det metaspråk som bedömarna använde i värderingen av språkfärdigheten kvalitativt sett. Det visade sig ändå vara svårt att omformulera lärarnas kommentarer till skriftliga beskrivningar. I det här skedet utelämnades beskrivningar som förhöll sig till elevens personlighet. För det andra (North 1996: 427-429) bads lärarna dela osorterade beskrivningar till olika kategorier. De beskrivningar som lärarna tolkade konsekvent togs med i den kvantitativa valideringen. Resultaten analyserades med en omarbetad klassisk ”*item banking*” metodologi, dvs. att resultaten på frågeformulären sorterades på basis av gemensamma drag i dem. Det gjordes också en DIF-analys (*differential item functioning*) (North 1996: 435) som analyserar variationen i beskrivningarnas tolkning i olika kontexter. De beskrivningar som inte fyllde DIF-analysens kriterier, utelämnades. Beskrivningarna för CEFR-skalor gjordes positiva med utgångspunkt i vad inläraren kan, och negativa formuleringar undveks. Som problem nämner North (1996) till exempel att det var svårt att göra beskrivningarna tillräckligt konkreta.

### **3.3 Utmaningar i CEFR-skalornas användning**

Vid användning av referensramen i bedömning av muntlig färdighet har man i ämnet nordiska språk vid Helsingfors universitet mött ett problem med referensramen. Enligt Green-Vänttinen och Rosenberg-Wolff (2007: 86-87) är skalorna mindre specifika ju högre upp man kommer. Studenterna borde befinna sig på C-nivån i slutskedet av studierna. Då är det svårare att bedöma när kriterierna inte är så specifika som på lägre nivåer.

I Kalliomäkis (2007) pro gradu-avhandling behandlas finska lärares åsikter om användningen av CEFR i evalueringsarbete. Kalliomäki (2007: 29, 35) intervjuade fem

språklärare i finska skolor om användningen av CEFR. Hennes resultat var att bara två av fem lärare använde skolskalan, eller CEFR, som skalan baserar sig på, i sitt arbete. Lärarna tyckte att skalornas uppmuntrande natur och kriterier för evalueringen i sig var bra, men de ansåg att skalornas bredd var problematisk och att nivåerna i skalorna liknar varandra mycket, vilket gör evalueringen svår.

En grupp tyska akademiker har kritiserat den europeiska referensramen hårt. Enligt Bausch, Christ och Königs (2003) har referensramen svaga teoretiska grunder. Den saknar en enhetlig teori som grund och användningen av terminologin är enligt skribenterna oklar och inkonsekvent. De anser också att referensramen är suspect när det gäller dess praktiska användning: den fokuserar på de kommersiella behoven hos utgivarna samt hos den internationella bedömarverksamheten som till exempel gäller medlemmarna i ALTE (*the Association of Language Testers in Europe*). Enligt kritikerna ignorerar referensramen den professionella litteraturen på tyska inom ämnet och använder bara litteratur på engelska och franska.

Starr Keddle (2004: 43) skriver att CEFR-skalan har både positiva och negativa sidor. Det finns utmaningar i hur den ska användas i skolor. Den tar inte hänsyn till grammatikbaserad utveckling i språkförmåga, vilket förorsakar en mur mellan beskrivningarna i skalan och inlärares prestationer. Enligt henne passar CEFR inte nödvändigtvis för inlärares i en skola, och den nödvändiga anpassningen förorsakar mera arbete. Också arbetet som uppstår i användningen och förståelsen av referensramen bildar hinder för dess vidsträckt användning i klassrummet. Enligt Starr Keddle (ibid. 50) finns det en allmän ton i referensramen som inte låter som den i klassrummet, vilket kan leda till att CEFR känns irrelevant för arbetet i klassrummet. Kriterierna handlar också om relativt mogna och vuxna teman och passar oftast inte samman med en tonårig inlärares intressen.

Hämäläinen et al (2007: 72) skriver att nivåskalorna i den europeiska referensramen är ett verktyg utan like för den finska språkundervisningen. Trots allt kan man ännu inte utnyttja nivåskalorna i full utsträckning på hela undervisningsfältet. De föreslår att



nivåskalorna skulle tas i bruk i praktiken inom planeringen, förverkligandet och evalueringen. Också Tarnanen et al. (2007) är av den åsikten att nivåskalorna borde tas i större utsträckning i bruk till exempel i studentskrivningarna i Finland. Kriteriebaserad evaluering skulle göra evalueringen mera genomskinlig och den skulle ge mera rörlighet inom det gemensamma europeiska utbildningsområdet (ibid. 394). Nu utnyttjas inte alla de möjligheter som CEFR erbjuder.

### 3.4 Den finländska nivåskalan för språkkunskaper

Den finländska nivåskalan för språkkunskaper används i den finländska grundskolan samt i gymnasierna. Skalan baserar sig på CEFR-skalan och används praktiskt taget på samma sätt som den. Den ingår i de nyaste läroplansgrunderna och finns på finska och på svenska.

I nedanstående tabell 2 jämförs nivåerna i de två skalorna. I nivåskalan har man använt en mera detaljerad nivåindelning än i den europeiska skalan. I CEFR finns det sex nivåer (levels) medan det i nivåskalan finns tio. A-nivåerna har man delat in i tre mindre enheter: A1.1, A1.2 och A1.3 i stället för A1 i CEFR. De andra nivåerna har indelats i två, till exempel B1 i B1.1 och B1.2. C2 finns inte med. (Grunderna för läroplanen 2004, i fortsättningen GFL 2004)

**Tabell 2: CEFR-skalan och nivåskalan för språkkunskaper.**

<b>CEFR-</b>	<b>Den finska nivåskalans nivåer</b>
A1	A1.1. De första grunderna i språket
	A1.2 Elementär språkfärdighet stadd i utveckling
	A1.3 Fungerande elementär språkfärdighet
A2 Waystage	A2.1 Den grundläggande språkfärdighetens
	A2.2 Grundläggande språkfärdighet stadd i utveckling

B1 Threshold	B1.1 Fungerande grundläggande språkfärdighet
	B1.2 Flytande grundläggande språkfärdighet
B2 Vantage	B2.1 Självständig språkfärdighet, grundnivå
	B2.2 Fungerande självständig språkfärdighet
C1 Effective	C1.1 Avancerad språkfärdighet, grundnivå
C2 Mastery	

(Eurooppalainen viitekehys 2003: 46, GFL 2004)

I *Grunderna för läroplanen i gymnasieutbildning för ungdomar* (2003) finns en tabell om de färdighetsnivåer i svenska som andra inhemska språket som eleverna borde uppnå i gymnasiet. När det gäller A-lärokurs, borde man finnas på nivå B1.2 (se tabell 3). I tabell 4 nedan finns exempelvis en förklaring av vad kunskapsnivån betyder i praktiken när det gäller talförmåga.

**Tabell 3: Färdighetsnivåerna i svenska som andra inhemska språket på gymnasienivå.**

Lärokurs x)	Hörförståelse	Tal	Läsförståelse	Skrift
A-lärokurs x)	B2.1 xx)	B1.2	B2.1	B1.2
B1-lärokurs	B1.2	B1.1	B1.2	B1.1

**Tabell 4: Färdighetsnivån på gymnasienivå, kriterierna för Tal på nivå B1.2.**

<p>Färdighetsnivån B1.2. Tal:</p> <ul style="list-style-type: none"> <li>• Kan med beskrivningar, specificeringar och jämförelser berätta om vanliga, konkreta ämnen och kan också redogöra för andra ämnen såsom filmer, böcker eller musik. Kan kommunicera säkert i de flesta vanliga situationer. Den språkliga framställningen är kanske inte särskilt exakt.</li> <li>• Kan uttrycka sig relativt obehindrat. Trots att pauser och avbrott förekommer fortsätter talet och eleven gör sig förstådd.</li> </ul>
--

- Uttalet är mycket lätt att förstå trots att intonationen och betoningen inte helt motsvarar målspråkets.
- Kan använda ett ganska omfattande ordförråd och vanliga idiom. Använder även många slags strukturer och komplicerade satser.
- Grammatikfel förekommer i viss mån men stör sällan ens mer omfattande kommunikation.

Kriterierna för språkfärdigheter i läroplansgrunderna är allmänna och inte språkspecifika. Arbetet med att koppla nivåskalan för språkkunskaper till CEFR har gjorts av Hildén och Takala (2007). De har testat kopplingar mellan CEFR och den finska nivåskalan med lärartest där ungefär 20 lärare ordnade osorterade beskrivningar enligt CEFR-nivåerna. Resultaten tyder på att lärarna var mycket eniga. Trots detta är förenligheten mellan de två skalorna endast nöjaktig, eftersom 65 % av de finska beskrivningarna kopplades tillbaka till deras korrekta ursprungliga nivåer i CEFR-skalan.

### **3.5 KORU-projektet inom de finska yrkeshögskolorna samt språkcenter vid finska universitet**

Åren 2003-2005 samarbetade de finska yrkeshögskolorna och språkcentren vid olika universitet i Finland för att göra bedömningen av språkkunskaper i svenska pålitligare, enhetligare och mera öppet (Elsinen & Juurakko-Paavola 2006). Målsättningarna i KORU-projektet var följande:

- 1) Huvudmålet var att förbättra högskolornas kommensurabelhet och genomskinlighet i evaluering av språkkunskaper i svenska  
Målet var också att
- 2) göra kriterierna och praxis för bedömningen av språkfärdigheter tydligare och enhetligare
- 3) utveckla svensklärarnas färdigheter i bedömning
- 4) bygga en ny och permanent verksamhetskultur som baserar sig på samarbete mellan och inom höskoleväsenden

I KORU-projektet deltog 25 yrkeshögskolor och 16 språkcenter vid olika universitet i Finland. I begynnelseskedet av projektet svarade svensklärarna på de deltagande högskolorna på en preliminär enkät. Resultatet blev att det finns variation i bedömningen mellan olika högskolor och mellan olika lärare. Betoningarna i bedömningen och

kriterierna varierade. Läsförståelse och ordförråd samt strukturer testades ofta, men få lärare berättade att de bedömde hörförståelse. Betoningen i bedömningen var ofta på skriftliga språkkunskaper, men mutliga språkkunskaper bedömdes också. Kriterierna för bedömningen hade man oftast gjort inom den egna högskolan med AMKKIA-skalan (AMKKIA = Kielitaidon arvioinnin yhdenmukaistaminen ammattikorkeakouluissa-projekt 2000-2001) som basis. Som ett resultat av KORU-projektet bearbetades det gemensamma bedömningskriterier för högskolesvenska. De förhåller sig till den europeiska referensramen enligt följande tabell:

**Tabell 5: Jämförelse av kravnivån i svenska vid högskolor i jämförelse med andra beskrivningar av språkkunskaper (Elsinen & Juurakko-Paavola 2006: 92). Min översättning.**

Statsförvaltningens språkexamen	Allmänna språkexamen	Den europeiska referensramen	Bedömning av kunskaper i svenska vid högskolor: beteckning i examensbetyg	Bedömning av kunskaper i svenska vid högskolor: kurs-/studieenhets-bedömning
Utmärkta kunskaper	Kunskapsnivå 6	Nivå C2	Kan inte bevisas i samband med studier i svenska vid högskola	Kan inte bevisas i samband med studier i svenska vid högskola
Goda kunskaper	Kunskapsnivåerna 4-5	Nivåerna B2-C1	Goda kunskaper	Vitsorden 4-5
Tillfredställande kunskaper	Kunskapsnivå 3	Nivå B1	Tillfredställande kunskaper	Vitsorden 1-3
	Kunskapsnivåerna 1-2	Nivåerna A1-A2		Underkänd

I tabell 5 ovan jämförs bedömningen av kunskaper i svenska på högskolenivå med nivåerna A1, A2, B1 och B2 samt C1 i den europeiska referensramen. A-nivån ger underkänd på högskolenivån. B1 är tillfredställande och goda kunskaper på högskolenivån motsvaras av nivåerna B2 och C1 i den europeiska referensramen. C2-nivån är enligt tabell inte aktuell på högskolenivån. I rapporten betonar skribenterna att bedömningen enligt KORU-nivåskalan är först och främst en holistisk bedömning som kan stödas med analytiska kriterier gällande olika delområden i språkkunskaper. När det gäller muntliga språkfärdigheter är de här delkriterierna kommunikationsförmåga,

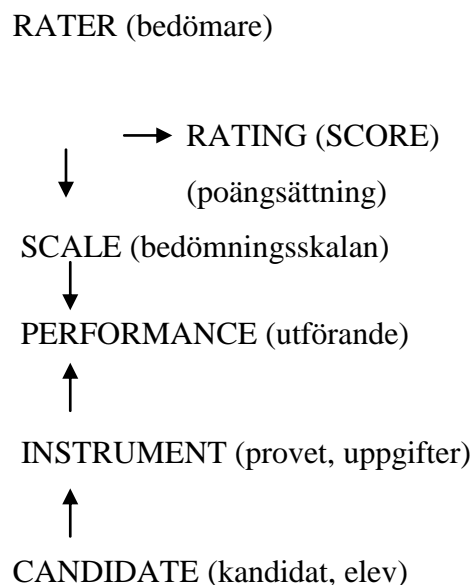
ordförråd, strukturer samt uttal. Det finns två olika nivåskalor: en för muntlig språkförmåga (*puhuminen*) och en för att skriva (*kirjoittaminen*).

## 4 Forskningsöversikt

I detta avsnitt redogör jag för forskningen om språkbedömning, som bedömningen av muntliga språkfärdigheter också hör till. För mitt arbete är forskning i muntliga språkprov och i bedömarbeteende väsentligast. Jag börjar med att beskriva allmänna principer i språkbedömningen och presenterar synpunkter på vad som gör ett prov effektivt och bra. Efter det tar jag upp olika aspekter i forskningen i bedömning av muntliga språkfärdigheter. Till sist koncentrerar jag mig på bedömaren. I viss mån hänger alla dessa delar ihop: ett prov är inte nyttigt utan en poängsättning och en bedömare kan inte bedöma utan en bedömningsskala. Reliabilitet och validitet är viktiga aspekter i språkbedömningen. Speciellt viktigt är det att fundera över reliabilitet när det gäller muntliga språkprov. Detta kom med också i resultaten som behandlas i kapitel 5.

### 4.1 Bedömaren i fokus

Relationerna mellan de olika aspekterna i språkbedömningen kan inledningsvis illustreras med en figur. McNamara (1996: 9) beskriver delfaktorerna i ett performanstest (*performance test*) med hjälp av följande figur.



**Figur 1: Performance-based assessment (McNamara 1996: 9). Min översättning.**

I figur 1 finns de olika delfaktorerna som påverkar ett muntligt prov. Begreppet performanstest förklaras noggrannare i avsnitt 4.4. Figuren ger en uppfattning om vilka faktorer som samverkar i ett prov av muntliga språkfärdigheter. Ett muntligt prov handlar oftast om ett utförande (*performance*). En kandidats utförande bedöms av en bedömare med hjälp av en skala. Resultatet blir en poängsättning (*score*). Min undersökning koncentrerar sig på bedömaren (*rater*) och de faktorer en bedömare samverkar med, dvs. skalan, kandidatutförandet och poängsättningen.

Ett tredje delområde som är väsentligt för min undersökning är CEF-skalan (2002) och dess tillämpning, nämligen skalan i de nyaste läroplansgrunderna (2002), som bedömarna i undersökningen använder som skala i sin bedömning. Skalorna presenterades i avsnitt 3.

## **4.2 Muntliga språkprov**

Som McNamara skriver (2000: 4) spelar olika slags prov en viktig roll i många liv:

”Language tests play a powerful role in many people’s lives, acting as gateways at important transitional moments in education, in employment, and in moving from one country to another.” Genom att delta i prov och bli godkänd får man ett betyg man kan gå vidare med, till exempel söka till en högskola eller kunna ansöka en arbetsplats. Prov är viktiga och därför måste de också vara så rättvisa och pålitliga som möjligt.

Språkprov är baserade på de teorier om språkanvändningen som provmakaren väljer. Synsättet kan vara till exempel kognitivt, interaktivt eller socialt. En forskare som har ett kognitivt synsätt betonar kunskaper och förmåga. Vid interaktiva och sociala synsätt är sociala roller och interaktion viktiga i provplaneringen.

*Data*, det språkliga materialet i en provsituation, tolkas enligt en teori så att man i det kan se bevis (*evidence*) på språkkunskaper. Man kan tolka informationen på olika sätt beroende på det synsätt man har som bakgrund i provet (ibid. 11). Om bedömarens slutsatser av den testades kunskaper är felaktiga, kan det ha allvarliga följder. Därför

måste provet valideras, dvs. riktigheten av de kunskaper som har satts som bas för provets utförande måste utredas till exempel med en kvantitativ analys.

Enligt McNamara (2000: 5-6) kan man dela in prov i två typer enligt metoden, traditionella språkprov med papper och penna samt performanstest, dvs. prov där språkförmåga evalueras i en kommunikationssituation. Prov simulerar beteende i det verkliga livet: "real world behaviour", och provutförandet evalueras som en indikator av hur personen skulle utföra en likadan uppgift i det verkliga livet (ibid. 8).

"Understanding testing involves recognizing a distinction between the *criterion* (relevant communicative behaviour in the target situation) and the *test*." Provet simulerar eller representerar det som kan hända utanför provsituationen, i det verkliga livet. Ett prov kan enligt McNamara (ibid.) inte ha som mätningssubjekt hur en person skulle bete sig i en autentisk situation, därför att man inte direkt kan observera det. I en provsituation är det av intresse hur den testade utför en uppgift. Muntliga prov är enligt honom ofta mer rollspelsaktiga än realistiska eller autentiska sampel. "Sometimes it is true that the materials and tasks in language tests can be relatively realistic but they can never be *real*." (ibid. 8) Den begränsade autenticiteten i ett språkprov förorsakar ett tomrum mellan provet och kriterierna. Detta i sin tur kan minska validiteten av provet. Frågan är om resultaten kan generaliseras. McNamara (ibid. 9) skriver också om observatörens paradox: "Observer's Paradox: that is, the very act of observation may change the behaviour being observed." När man blir evaluerad, är situationen ofta stressande och man kan bli nervös. Det kan också vara lättare att samtala med kameran än med en riktig människa.

Brown (2004: 3) definierar ett prov som en metod att mäta en persons förmåga, kunskap eller utförande i en viss domän. Ett prov måste vara en metod, ett instrument med vilket man mäter en individ. Ett prov mäter individens kunskaper eller utförande. Ett språkprov mäter kompetens, med andra ord språkkunskaper, i en viss domän. Brown (ibid. 3) skriver att "performance-based tests sample the test-taker's actual use of language, but from those samples the test administrator infers general competence". Enligt honom (ibid. 11) innebär performansbaserade prov mera direkt mätning eftersom studerandena

evalueras när de utför verkliga eller simulerade uppgifter. Tekniskt taget högre innehållsmässiga validitet är nådd eftersom studerandena evalueras när de utför de målsatta lingvistiska handlingarna. Brown är alltså av den åsikten att inlärarna i performansbaserad evaluering utför det beteende som man vill evaluera. I interaktiva uppgifter evalueras inlärarna när de talar, frågar och svarar. Enligt honom kan ett bra prov nå samma nivå av autenticitet som språkanvändningen i det verkliga livet.

Enligt den europeiska referensramen (Eurooppalainen viitekehys 2003: 256) kan kompetens aldrig testas direkt. Man måste stödja sig på bevis på prestationsförmågan och göra generaliseringar av den testades språkförmåga med hjälp av dem. Om man mäter prestationsförmågan i praktiken, är det oftast fråga om ett performanstest (*suoritustesti, performance test*). Det betyder en prestation i en nästan autentisk situation. Man kan tolka muntliga språkprov som performanstest därför att tolkningarna av de testades språkförmåga baserar sig på bevis på hur de testade behärskar väsentliga diskurstyper i inlärningskontexten.

För att summera kan man konstatera att McNamara (2000), Brown (2004) och den europeiska referensramen (CEFR 2001) förhåller sig alla lite olika till validiteten i språkprov. Enligt McNamara kan man inte mäta autentiskt språkbruk med prov, eftersom provsituationen inte kan vara autentisk eller äkta kommunikation. Enligt Brown kan ett bra konstruerat prov nå samma autenticitetsnivå som språkanvändningen i det verkliga livet. I den europeiska referensramen förhåller man sig lite försiktigare till saken, men det konstateras, att situationen i performanstest är åtminstone nästan autentisk.

Luoma (2004: 153) tar upp rollspelsuppgifter som ett exempel på hur man kan göra kommunikationen i ett prov mångsidigare. Rollspelen gör det möjligt att de testade tar en ny roll och ett nytt, simulerat förhållande till kommunikationsparten. Om rollen och situationen är väsentlig för provets syfte, får bedömare nyttig information och ett nytt perspektiv till den testades språkkunskaper. Simulerade uppgifter liknar rollspel, eftersom deltagarna i dem agerar i en fiktiv kommunikationssituation. Rollspel och



simulering kan också förorsaka problem i proven, därför att skådespelarprestationen kan påverka utförandet av uppgiften och leda till orättvis bedömning.

Performanstestens användning i bedömningen av muntliga språkfärdigheter ökar kontinuerligt, och performanstest används också i yrkesinriktad språkevaluering (Brown & Lumley 1997: 137). Enligt författarna kan deltagarna i performanstest i större utsträckning visa hur de kan använda målspråket än vad de vet om det. Detta hänger ihop med den nuvarande synen på språket vilken betonar kommunikativ kunskap. Utförandets förutsäggande validitet är sannolikt högre i performanstest när kandidaterna deltar i simulation av relevanta utmaningar i kommunikationen på arbetsplatsen.

### 4.3 Bedömare

Alderson et al. (1995: 105) förklarar att en bedömare är den person som är ansvarig för bedömning av en kandidats utförande i ett prov eller examen. De skiljer mellan bedömare och interlokutör i bedömningen av muntligt språk. Interlokutör är i interaktion med den testade kandidaten medan bedömaren evaluerar kandidatens uppförande.

Alan Davies et al. (1999: 161) definierar *rater* dvs. på svenska bedömaren – det finns ingen definition på *examiner* - som följer.

*rater*

*Also judge, marker, scorer*

*The judge or observer who operates a **rating scale** in the **measurement of oral and written proficiency**. The **reliability** of raters depends in part on the quality of their training, the purpose of which is to ensure a high degree of comparability, both inter- and intra-rater. Since raters are human and are therefore subject to individual **biases**, close attention is paid not only to reliability, but also to analyses of rater bias.*

Författarna (ibid.) definierar bedömaren noggrannare än Alderson et al. (1995). Enligt Davies et al. (1999) opererar bedömaren med en bedömningsskala som mäter muntlig och skriftlig färdighet. De tar också upp reliabilitetsfrågan: en bedömares reliabilitet, dvs. bedömningens pålitlighet (se avsnitt 4.6.2), påverkas av personens utbildning samt

individuella avvikelser (*bias*). Evalueringen förklaras som en allmännare term som betyder okontrollerad vardaglig observation.

I den europeiska referensramen (2003: 242) definieras bedömning och evaluering också som två olika begrepp. Bland annat språkprov räknas som bedömning (*arviointi*), medan evaluering (*evaluointi*) uppfattas som en bredare term. Man kan evaluera många andra faktorer också än bara inlärarens språkkunskaper. Man kan till exempel evaluera material, diskursens kvalitet eller undervisningens effektivitet.

Bedömaren har en viss roll i bedömningsprocessen. Speciellt när det gäller bedömning av muntliga språkfärdigheter, blir bedömaren ofta mera synlig i bedömningsprocessen. Då är det inte fråga om att korrigera blanketter där oftast bara ett svar är korrekt, utan bedömningen kan vara mera krävande. Enligt McNamara (1996: 117) omfattar performansbedömning nödvändigtvis subjektivt omdöme. Det kan inte göras automatiskt. Värdefulla bedömningar är komplexa och innehåller nödvändigtvis bedömarens tolkning. Därför orsakar de också oenighet. McNamara nämner noggrann utbildning och kontrollering av bedömningen som verktyg till att minska oenigheten mellan bedömarna. Grova skalor, som till exempel bara numrering som skala, försämrar reliabiliteten. Enligt McNamara (*ibid.*) räcker det inte bara med utbildning och kontrollering, eftersom de inte eliminerar oenigheten totalt.

#### **4.4 Bedömning av muntlig språkfärdighet**

Bedömning av muntlig färdighet är en mångfacetterad process. Processen innehåller många nivåer, och varje nivå agerar och interagerar så att de producerar någonting till följande nivå. Luoma (2004: 5) definierar bedömningen av muntliga språkfärdigheter som en cykel som förenklat beskrivet innehåller fyra aktiviteter. Först uppstår ett behov för bedömning av muntliga språkfärdigheter, vilket leder till planering och utveckling. Detta innehåller utveckling av evalueringskriterier, uppgifter och administrativa procedurer. Efter det kan bedömningssättet börja användas. De två sista processerna är interaktiva: genomförandet av provet, då de testade interagerar och visar ett exempel eller sampel på sina kunskaper, samt bedömningen eller evalueringen, då bedömarna

använder kriterierna för utförandet av uppgifter. Resultatet blir poäng eller muntlig respons. McNamara (2000: 23) beskriver bedömningen också som en cykel som genererar evidens om sina egenskaper. När planerarna utvecklar ett nytt prov och dess kriterier, måste de ta hänsyn till många restriktioner, till exempel om det finns tillräckligt med tillgängliga utbildade bedömare. I nästa skede börjar man planera provets domän och innehåll samt metod.

Luoma (2004: 1) skriver att bedömning av muntliga språkfärdigheter är utmanande, eftersom det finns så många faktorer som påverkar vårt intryck av hur bra en person talar ett språk och eftersom vi förväntar oss att poängsättningen är exakt och rättvis.

Enligt Tim McNamara (2000: 10) är bedömare i en provsituation inte intresserade av det observerade språkbruket i sig, utan mera av vad detta utförande berättar om potentiella senare utföranden i samma slags situationer utanför testsituationen, dvs. hur man klarar av kommunikativa situationer i det verkliga livet. Man försöker i utförandet i testsituationen se de egenskaper som ligger bakom det.

Då man jämför den europeiska referensramens förhållningssätt med McNamaras (2000), kommer man till att de liknar varandra. Enligt referensramen kan man aldrig mäta kompetens direkt. Man måste lita på sampel som visar prestationsförmåga och försöka göra generaliseringar om språkförmågan på basis av dem. Kompetens finns endast i språkbruket. I alla test bedöms bara utförandet, fastän man kan försöka dra slutsatser på basis av materialet om hurdan kompetens som finns i bakgrunden. (Eurooppalainen viitekehys 2003: 255)

## **4.5 Teoretiska modeller för språkbedömning**

McNamara (2000) utreder språkevalueringens historia och olika teoretiska synpunkter som används som bas när språkprov konstrueras. McNamara tar upp kommunikativa modeller. Enligt honom har synsätten nu och tidigare för det mesta koncentrerat sig antingen på kognitiva aspekter eller på språkbrukets aspekter i provsituationer (ibid. 21). Luoma (2004: 96) konstaterar, att teorier inte alltid används som bas för ett språkprov.

Också till exempel läroplansgrunderna kan fungera som referens för provet. Läroplansgrunderna kan basera sig på en teoretisk ram eller en kombination av många olika synsätt på språkinläringen.

Eklund Heinonen (2009: 44) delar testteorin historiskt sett in i tre perioder: den förvetenskapliga perioden, den psykometrisk-strukturalistiska perioden under 1930-1960-talen och den psykolingvistisk-sociolingvistiska perioden under 1960- och 1970-talen. Dessa tre perioder föregick den kommunikativa synen på språkfärdigheter som enligt Eklund Heinonen (2009) råder nuförtiden.

Under den förvetenskapliga perioden var man inte orolig över till exempel reliabilitet eller annan objektivitet utan litade på forskarnas expertis. Under den psykometrisk-strukturalistiska perioden grundade sig, enligt Eklund Heinonen (2009: 44,) på behavioristisk psykologi samt strukturalistisk syn på språk. Enligt den behavioristiska inläringsteorin är språket som ett inlärt beteende, och inläring av ett nytt språk påverkas av tidigare beteendemönster, till exempel ett redan inlärt förstaspråk. Strukturalismen koncentrerade sig på att analysera och systematisera språkinläringen och jämförde strukturer i inlärares första- och andraspråk. Också enligt McNamara (2000: 13) baserade sig tidiga teorier i språkevalueringen på en strukturalistisk syn, dvs. språkkunskaper sågs som kunskaper av språket som ett system. Detta betydde att grammatik och ordförråd betonades i evalueringen och ofta gjordes detta genom dekontextualisering, så att man tog textutdrag ur sin kontext och använde dem i tester utan någon kontextram, och så att olika delar av kunskap testades isolerade från varandra. Eklund Heinonen nämner enfokusuppgifter och flervälsfrågor som exempel på typiska testuppgifter under den strukturalistiska tiden. Man ansåg att fri muntlig och skriftlig produktion inte kunde mätas på ett objektivt sätt, och därför undveks dessa typer av uppgifter i språktester.

Senare, i början av 1970-talet, uppstod ett behov för evaluering av praktiska språkfärdigheter och kommunikation. Betoningen flyttades från den formella kunskapen till ett mera kommunikativt sätt att använda språket - i en kontext. Den psykolingvistisk-

sociolingvistiska perioden ledde till den kommunikativa språksynen (Eklund Heinonen 2009: 45). Senare på 1970-talet konstaterade en amerikansk forskare, John Oller, att *cloze tests*, lucktester mäter ”the ability to integrate grammatical, lexical, contextual and pragmatic knowledge in test performance.” (McNamara 2000: 15). Hymes teori om kommunikativ kompetens började samtidigt påverka språkundervisningen från och med 1970-talet och tio år senare började man planera kommunikativa språkprov (*communicative language tests*). Då började språket ses som ett socialt fenomen, inte längre bara som ett psykologiskt. Autentiska texter och uppgifter från det verkliga livet började användas i språkprov, först i Britannien och senare annanstans. McNamara (2000: 16-17) definierar kommunikativa språkprov med två kännetecken. För det första var de performansprov som krävde bedömning då inläraren utförde en vidsträckt kommunikativ handling, receptiv, produktiv eller både och. För det andra betonade de de sociala roller som kandidaterna skulle inta i en verklig situation samt erbjöd verktyg för detaljerad specificering av kraven för sådana roller. Enligt Eklund Heinonen (2009) började man under 1980-talet fokusera alltmer på inlärarens förmåga att använda språket i verkliga situationer. Hymes protesterade mot Chomskys teorier som betonade grammatisk kompetens. Hymes tyckte att språklig kompetens innehåller flera olika kompetenser varav den grammatiska kompetensen är bara en. Språkinläraren måste också lära sig använda språket lämpligt i ett visst sammanhang.

Kommunikativ kompetens, som Hymes introducerade och andra sedan dess har utvecklat, betonar språkets användare och användningen av språket för kommunikation (Luoma 2004: 97). Andra typer av språkprov har ändå inte försvunnit, utan scenen för språkbedömningen ser för tillfället ganska eklektisk och mångsidig ut (McNamara 2000: 21).

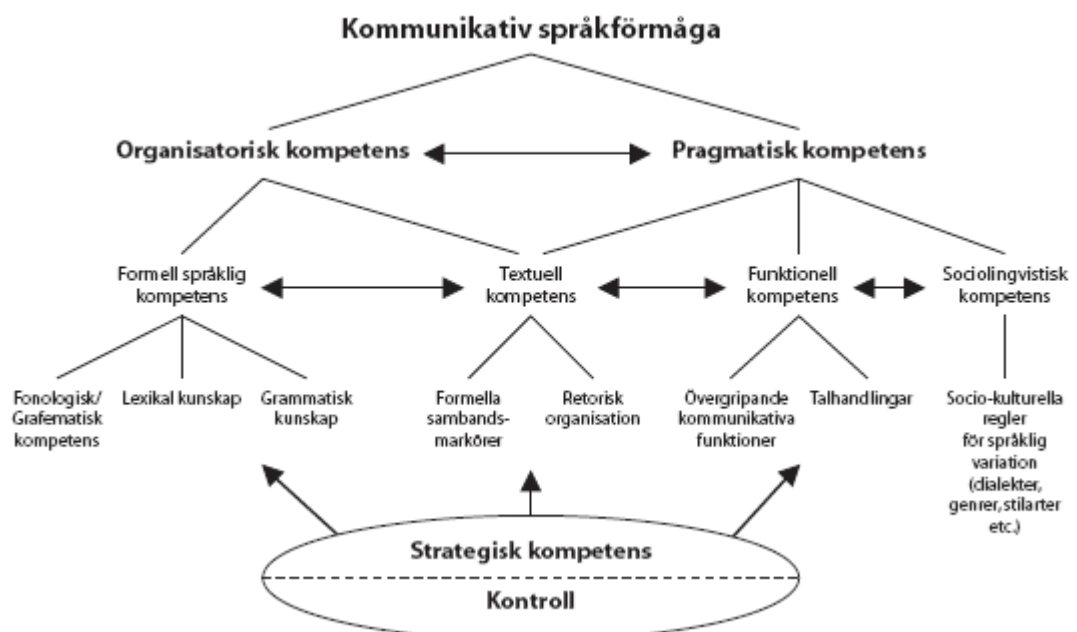
McNamara (ibid.) skriver, att det har forskats i vilka andra faktorer än kunskap som påverkar utförandet av uppgiften i ett språkprov. Sådana faktorer är till exempel självförtroende, motivation och emotionellt tillstånd. Man är medveten om dessa faktorerers komplexitet, vilket gör det mera komplicerat att försöka standardisera villkoren för evalueringen för att kunna göra den så rättvis som möjligt. Det finns problem med att

kunna nå en rättvis evaluering av performanstest. Forskarna har tänkt att den kommunikativa språkförmågan består av olika delar, men enligt McNamara (ibid. 20) är detta ett problematiskt synsätt därför att de olika komponenterna har analyserats icke relaterade till varandra: "There is still therefore the problem, which models of communicative competence were designed to resolve, of how to account for the way the different aspects act upon each other in actual communication." McNamara (1996) konstaterar, att Hymes förståelse av kommunikativ kompetens var vid och utvecklad i en L1-kontext, dvs. den syftade på förstaspråksanvändare. Kommunikativ kompetens i en L2-kontext tenderar att uppfattas som smalare än det hymesiska konceptet. Den tar inte med sociokulturella faktorer utan koncentrerar sig på att evaluera språkförmåga. McNamara (ibid. 85) konstaterar, att vid ett muntligt prov är kandidaten i social interaktion med intervjuaren eller en annan person. Forskningen borde således koncentrera sig mera på social interaktion och sociolingvistik i en L2-provsituation.

Enligt Eklund Heinonen (2009: 46-49) har Bachmans modell om kommunikativ språkförmåga haft stor inverkan på språktestning under den senaste tiden. Bachman och Palmer (1996) har vidareutvecklat Bachmans (1990) modell. Modellen är avsedd att användas vid konstruktion och tolkning av muntliga språktest. För att kunna dra slutsatser om en persons språkförmåga, måste man definiera vad som menas med språkförmåga och skilja språkförmågan från andra faktorer i ett test som kan påverka testtagarens resultat. Definitionen av språkförmåga ligger till grund för de slutsatser man kan dra på basis av testtagarens prestationer i en testsituation. Prestationer kallas ett konstrukt som betyder de egenskaper som språkförmågan baserar sig på och som man försöker mäta i ett språktest. Nuförtiden anses det att språkbehärsknigen är någonting komplext. Inläraren måste lära sig använda språket effektivt i olika kommunikativa situationer, vilket kräver olika kunskaper på många plan. Kommunikativ språkförmåga innehåller enligt Bachman & Palmer (1996, genom Eklund Heinonen 2009) två huvudkategorier: organisatorisk kompetens och pragmatisk kompetens (översättning av: Eklund Heinonen 2009). Organisatorisk kompetens (se figur 2 nedan) innehåller tolkning och producering av grammatiskt korrekta yttranden och deras sammanfogande till längre texter eller talade enheter. Denna kompetens innehåller kunskaper om språkets ljudsystem, ordförråd och grammatik. Bachman och Palmer delar den organisatoriska

kompetensen vidare i formell språklig kompetens och textuell kompetens. Den formella kompetensen innehåller kunskap om formella byggstenar i språket: kännedom om till exempel fonologi, syntax, stavningsregler osv. Textuell kompetens däremot betyder förmågan att av korta yttranden kunna skapa längre förståeliga och koherenta talade och skrivna olika typers texter.

Pragmatisk kompetens i Bachman och Palmers modell (se figur 2 nedan) betyder å sin sida kunskap om språkanvändningsregler (ibid.). Språkinläraren skall kunna tolka och producera yttranden lämpliga för olika situationer och för olika kommunikativa funktioner. Pragmatisk kompetens delas in i två delfaktorer: funktionell kompetens och sociolingvistisk kompetens. Eklund Heinonen (2009: 47) förklarar den funktionella kompetensen som en förmåga att effektivt ”uttrycka sina kommunikativa avsikter med hjälp av olika språkliga funktioner”. Dessa funktioner är till exempel informerande, expressiva, direktiva och kontaktreglerande. Den sociolingvistiska kompetensen innehåller för sin del att man kan anpassa och tolka språket enligt situationen och samtalspartnern. Det täcker också regler om dialekter, register och olika genrer samt om idiomatiska och metaforiska uttryck.



**Figur 2: Modell över den kommunikativa språkförmågan (efter Eklund Heinonen 2009: 48)**

Nederst i figur 2 ovan finns strategisk kompetens som betyder mer allmänna kognitiva kunskaper för att språkinläraren skall kunna utnyttja sina språkliga resurser optimalt (Eklund Heinonen 2009:48). Strategisk kompetens betyder en kognitiv problemlösningsförmåga som kan användas till att lösa antingen språkliga problem eller andra problem. Den strategiska kompetensen behövs för att språkanvändaren skall kunna använda sin språkliga kompetens i naturlig kommunikation. Dessutom tas kontroll med i modellen vilket Eklund Heinonen beskriver. Kontrollen har introducerats av Lindberg (1997, genom Eklund Heinonen 2009: 49). Kontroll innebär att språkkunskaperna är automatiserade så att språket är flytande och ingen större ansträngning behövs. Graden av kontroll kan variera under inlärningsprocessens olika stadier, men också mellan olika kommunikationssituationer. Till exempel nervositet, stress och trötthet kan påverka kontrollen negativt. Detta kan vara viktigt speciellt i stressade testsammanhang. Eklund Heinonen konstaterar också att fastän de ovan presenterade komponenterna i den kommunikativa språkförmågan känns väl uppdelade och avgränsade, är språkförmågan dynamisk. De olika delkompetenserna samverkar med varandra och påverkas också av andra mentala processer som till exempel av kunskaper om omvärlden, känslor och värderingar samt av språkanvändarens personliga egenskaper (ålder, kön osv.). Man kan inte heller glömma kommunikationssituationens och samtalspartnernas betydelse för hurdan språkanvändningen blir.

Den kommunikativa synen på språkförmågan kan kännas komplex och abstrakt, men det har haft en tydlig inverkan på praktiken. Enligt McNamara (2000) borde man ändå börja se utförandet inte bara som en kollektion av olika särdrag, utan mera som en integrerad helhet där särdragen samspelar. Man borde också börja se kandidatens utförande som en del av en komplex konstruktion som innehåller deltagarna i samtalet, provmakarna och bedömarna. Nu betraktas den testade enligt honom som isolerad av andra i provsituationen. McNamara (2000) betonar också provutförandets sociala karaktär. Den kommunikativa synen blir enligt min åsikt speciellt synlig vid bedömning av muntliga språkfärdigheter. I mitt material är testuppgifterna kommunikativa, och bedömarna måste ta kommunikationen i beaktande också i bedömningen.



## 4.6 Principer i språkbedömningen

I den här delen skall jag introducera de olika principerna i språkbedömningen. Bedömningen av muntliga språkfärdigheter är mångfacetterad och det finns många saker man måste ta hänsyn till samtidigt. Också olika språkvetare håller sig olika till språkbedömningens principer. Jag skall här presentera Browns (2004) och Bachman & Palmers (1996) syn på språkbedömning. Andra intressanta teoretiker är McNamara (2000) samt Luoma (2004) som jag också skall ta upp. Till sist presenterar jag synen i CEFR på språkbedömning.

Kvaliteten i ett prov är ett viktigt diskussionsämne i intervjumaterialet. Därför vill jag koncentrera mig på att diskutera forskarnas syn på kvaliteten i språkprov. I det följande redogörs speciellt för vad som gör ett prov effektivt och bra enligt de ovannämnda teoretikerna.

Bachman & Palmer (1996: 18) anser, att en tests användbarhet bör beaktas när språktester planeras. I användbarheten tillhör reliabilitet, begreppsvaliditet, autenticitet, interaktivitet, testeffekt och genomförbarhet (översättningarna av Eklund Heinonen 2009: 49). Dessa egenskaper inverkar tillsammans på den övergripande användbarheten av ett prov.

Brown (2004) listar fem principer i språkevaluering. Med dem kan man ”testa tester”, med andra ord testa om ett prov är effektivt. Principerna är validitet, reliabilitet, autenticitet, genomförbarhet och washback.

Enligt Eklund Heinonen (2009: 50) har man flyttat betoningen från 1980-talets kvantitativa forskning av muntliga språktester till en mera kommunikativ syn under 1990-talet. I stället för korrelationskoefficienter och andra kvantitativa statistiska mått har man börjat studera muntliga tester som kommunikationssituationer, och forskat hur kommunikationen påverkar testens reliabilitet och validitet. Därmed har man blivit mera intresserad av själva den språkliga produkten vid muntliga test istället för att fokusera på

testresultat när man testar reliabilitet och validitet (Brown, A. 2004: 254 I: Eklund Heinonen 2009: 51).

#### **4.6.1 Bedömningens validitet**

Enligt Brown (2004) är validiteten den viktigaste principen i bedömningen. Validitet betyder att provet testar det som det borde testa, till exempel att ett prov som mäter läsningsförmåga verkligen mäter läsningsförmåga. Brown nämner fem olika aspekter av validiteten som kan ses som bevis på att validiteten är bra i ett prov. Aspekterna kallas *content-related evidence*, *criterion-related evidence*, *construct-related evidence*, *consequential validity* och *face validity*.

Davies et al. (1999: 221) beskriver validitet på följande sätt. Ett språkprov är valitt om det bildar en exakt, konkret representation av ett abstrakt begrepp som till exempel kunskap. De vanligaste validitetstyperna är enligt författarna (egen översättning) innehållsvaliditet (*content validity*), begreppsvaliditet (*construct validity*), samverkande validitet (*concurrent validity*), förutsäggande validitet (*predictive validity*) och ytvaliditet (*face validity*). Innehållsvaliditet mäter om innehållet i provuppgifterna är lämpligt och tillräckligt mångsidigt för att bedömning ska kunna utföras. Med begreppsvaliditet kan det mätas om provet representerar den språkinlärningsteori den borde representera. Samverkande validitet mäter relationen mellan vad som mäts i ett prov och i ett annat kriterieredskap. Förutsäggande validitet mäter hur bra ett prov förutsäger utförandet med externa kriterier. Med detta avses att ett provresultat korrelerar med resultaten före eller efter provet i andra sammanhang, till exempel på en kurs. Ytvaliditet betyder att ett prov verkar mäta den kunskap eller den förmåga som den påstår mäta ur en utbildad iakttagares synvinkel.

I alla aktiviteter i bedömningscykeln försöker man försäkra kvaliteten av bedömningen. Luoma (2004) nämner *construct validity* (s. 7), begreppsvaliditet som ett exempel på kvalitetsförsäkringen. Begreppsvaliditet betyder att man vill försäkra sig om att man bedömer rätt sak i provet. Enligt Luoma är begreppsvaliditeten den viktigaste egenskapen i all bedömning. Brown (2004) nämner färdighet och kommunikativ

kompetens som exempel på lingvistiska begrepp (*construct*), och självkänsla och motivation som psykologiska begrepp. Enligt honom betyder begreppsvaliditet att provet behandlar de teoretiska begrepp som man har bestämt att det ska behandla.

#### **4.6.2 Bedömningens reliabilitet**

Reliabilitet betyder att ett prov ger samma resultat oberoende av situationen och de testade. Detta kan påverkas av till exempel studentrelaterade orsaker som ångest eller tillfällig sjukdom samt *test-wisness*, provkloket eller skicklighet i att ta prov. Brown (ibid.) nämner också evalueringens reliabilitet, den administrativa reliabiliteten och provets reliabilitet.

Reliabilitet betyder enligt Bachman & Palmer (1996) hur pålitlig och konsistent mätningen är. Inkonsistens är variationen i testpoäng som beror på andra faktorer än de konstruktioner som man vill mäta. Omotiverad variation kan minimeras med testdesign så att testuppgifterna mäter de drag som avsetts. Begreppsvaliditet hänför sig till att de tolkningar som görs på basis av testpoäng är meningsfulla och lämpliga.

Davies et al (1999: 168-169) förklarar reliabiliteten som den verkliga nivån av överensstämmelse mellan ett resultat i ett prov och ett annat resultat i samma prov eller i ett olik prov. Idealiskt vore om det inte fanns olikheter mellan resultaten och om de stämde överens, men i praktiken finns det mättningsfel i test. De kan förorsakas av avvikelser (*bias*) i val av poster (av vad bedömningen görs) i provet, av avvikelser i testtid eller av bedömaravvikelser. När reliabiliteten av ett objektiva prov värderas, gäller värderingen oftast de två förstnämnda typerna av avvikelser. I ett subjektivt prov, som alltså bedöms subjektivt, är bedömaravvikelsena de viktigaste som man måste ta hänsyn till. Vanligen ses reliabilitet som en nödvändig men inte tillräcklig egenskap i ett prov. Reliabilitet och validitet borde hänga tätt ihop. Reliabiliteten fokuserar på empiriska aspekter i bedömningsprocessen då validiteten betonar de teoretiska aspekterna i processen. Enligt Hildén (2008: 2) har man tidigare försökt mäta validitet via reliabilitet, eftersom reliabiliteten antogs vara en nödvändig förutsättning för validiteten. Senare

konstaterades det att reliabiliteten inte nödvändigtvis är en tillräcklig förutsättning för validiteten.

#### **4.6.3 Interbedömarreliabilitet och intrabedömarreliabilitet**

Reliabiliteten mellan bedömare (interbedömarreliabilitet enligt Eklund Heinonen 2009: 53) betyder graden av konsensus i två eller flera självständiga bedömares bedömning av en kandidats utförande. Detta är speciellt viktigt att ta hänsyn till i test som bedöms subjektivt. (Davies et al. 1999: 88.) Om bedömningen varierar kan det ha allvarliga följder för de testade kandidaterna. Därför är det viktigt att ha så bra reliabilitet mellan bedömare som möjligt. Ofta är det så, att enigheten mellan bedömare är större i de extrema fallen i båda ändorna på bedömningsskalan, men oenigheten ökar i mitten av skalan. Davies et al (ibid.) föreslår, att i rättvisans namn borde subjektivt bedömda test ha två eller flera bedömare om testet är betydelsefullt. Ju flera bedömare, desto mindre risk finns det för fel i kandidaternas poängtal. Skillnaderna i bedömningen mellan bedömare blir problematiska om olika bedömare kommer fram till olika resultat vid bedömningen av samma prestation när skillnaderna inte har något med testtagarens språkkunskaper att göra, utan bedömare tolkar bedömningskriterierna på olika sätt (Wrigstad 2004: 720 f. genom Eklund Heinonen 2009: 49). Bedömare kan också inverka på testets validitet och sambandet mellan de kriterier som bedömare använder i sin bedömning och de som uttrycks i bedömningsunderlaget. Ett test kan avses bedöma både pragmatisk och organisatorisk kompetens om båda kompetenser täcks av kriterierna som används i bedömningen, men det kan ändå hända att bedömare enbart fokuserar på den ena aspekten i sin bedömning (Carlsen 2003:90 genom Eklund Heinonen 2009:53). Detta minskar testets validitet eftersom olika bedömare fokuserar på olika saker i bedömningen. På detta sätt kan interbedömarreliabiliteten vara hög utan att validiteten i bedömningen är hög.

Reliabiliteten hos en bedömare betyder graden av konsistens i en bedömares subjektiva bedömningsarbete, det vill säga om bedömningsresultaten är likadana i samma slags testsituationer. Man kan mäta konsistensen genom att jämföra två självständiga

bedömningar av samma testutförande, båda gjorda av en och samma bedömare. (Davies et al 1999: 91)

#### 4.6.4 Autenticitet

Med autenticitet avses i ett språkprov att uppgiften sannolikt förekommer också i det verkliga livet. ”Essentially, when you make a claim for authenticity in a test task, you are saying that this task is likely to be enacted in the ’real world’.” (Brown 2004: 28.) Autenticiteten kan vara synlig till exempel i:

- språket i provet är så naturligt som möjligt.
- uppgifterna är kontextualiserade, inte isolerade.
- temana är meningsfulla, relevanta och intressanta för inläraren.
- temana är tematiskt organiserade, till exempel i form av story line eller episod.
- uppgifterna står för uppgifter i det verkliga livet.

Brown (ibid.) skriver att provuppgifternas autenticitet har ökat under de senaste åren. Tidigare, för två eller tre årtionden sedan var provuppgifterna enligt honom osammanhängande och långtråkiga. Nuförtiden använder man mera autentiska källor för textavsnitt till exempel i läsförståelseuppgifter.

Enligt Bachman & Palmer (1996: 39) är autenticitet ett viktigt drag i språkprov av två orsaker. För det första utgör den en länk mellan testutförandet och de TLU (*target language use*) domäner som testets resultat generaliseras till. För det andra kan sättet med vilket testtagarna uppfattar testuppgiftens relativa autenticitet potentiellt främja deras testutförande.

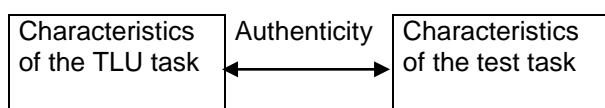
Bachman & Palmer (1996) definierar autenticitet som graden av motsvarighet mellan drag i en viss uppgift i ett språkprov och drag i en TLU-uppgift. Interaktivitet betyder graden av i vilken utsträckning de kompetenser som man vill bedöma är med i genomförandet av uppgiften. De mest relevanta kompetenserna för språkbedömning är

kandidatens språkförmåga, vilket innehåller språkkunskaper och strategisk eller metakognitiv kompetens, samt ämneskunskap (*topical knowledge*) och emotionellt schema. Ämneskunskap betyder i vid mening kunskaperna i långtidsminnet, dvs. all den information om världen som man kan referera till när man använder språket. Emotionellt schema betyder affektiva eller emotionella korrelater i ämneskunskap. I en provsituation betyder ett emotionellt schema den emotionella reaktionen till provuppgiften. Om uppgiften är emotionellt laddad (handlar t.ex. om abort), kan det antingen främja eller begränsa kandidatens språkbruk i provsituationen. Emotionella reaktioner kan vara positiva eller negativa och påverka provutförandet både positivt och negativt. Bedömarna borde vara medvetna om detta (ibid. 65). Interaktivitet betyder alltså hur mycket och på vilket sätt testtagarna deltar i genomförandet av uppgiften med sin språkförmåga, ämneskunskap och emotionella inblandning. Bachman & Palmers definition av interaktivitet i en testsituation avviker från vad man i språkvetenskap allmänt avser med interaktivitet i ett samtal.

Både autenticitet och interaktivitet är relativa begrepp, och man borde tala om mera eller mindre autentiska uppgifter, inte antingen autentiska eller icke-autentiska. En uppgifts autenticitet och interaktivitet kan definieras så att de jämförs med testtagarnas särdrag, med TLU-domänen och med testuppgiften. (ibid. 39.)

Lewkowicz (1997: 166) skriver i sin artikel om autenticitet att autenticitetsfrågan har diskuterats från och med 1970-talet, då ett krav på autenticitet uppstod i språkundervisningen och evalueringen samtidigt med kommunikativa metoder (ibid.). Enligt Lewkowicz är autenticitet ett problematiskt begrepp eftersom samförstånd om vad det innehåller inte finns. En del forskare betonar dess betydelse mera än andra. På 1970-talet började man betona autenticitet av testinflöde, dvs. av det material som användes i prov. Texterna borde inte vara förenklade för testsituationen, utan borde användas som sådana. Då tog man inte hänsyn till att en text är producerad för ett speciellt kommunikativt syfte och i en viss social situation. I en annan situation kan texten bli tolkad på helt olika sätt än i originalkontexten. Senare började man modifiera texter för provändamål.

Uppgifter som avspeglar språkliga handlingar i det verkliga livet har också setts som problematiska, eftersom de alltid är uppgifter med bedömningssyfte, och därför kan de inte avspegla det verkliga livet (ibid. 167). Bachman & Palmer (1996: 9) tar med autenticiteten som en egenskap som mäter ett provs användbarhet. Enligt författarna innebär autenticitet i hur hög grad kännetecknen i en viss uppgift i ett språkprov motsvarar språkbruket i målspråket, dvs. en uppgift i det verkliga livet. I figur 3 nedan visas relationen mellan en uppgift i målspråket och en uppgift i ett språkprov. Bachman & Palmer (1996) anser, att det är viktigt att en uppgift i ett språkprov motsvarar språkbruket i verkligheten för att man ska kunna motivera användningen av språkprov.



**Figur 3: Autenticitet (ibid. 1996: 23)**

Lewkowicz (1997: 182) konstaterar, att det är svårt för lärarna att särskilja autentiska texter från texter som är simulerade för provändamål samt att för testtagarna spelar autenticiteten inte en betydande roll i provsituationen. Hon ifrågasätter autenticitetens betoning som en viktig del av språkbedömningen, eftersom autenticiteten verkar vara viktig endast i teorin och inte ha en stor betydelse för andra intressenter i testprocessen.

## **4.7 Andra faktorer i språkbedömningen**

*Washback* betyder de effekter som evalueringen har på undervisningen och på inläringen. Brown (2004) förklarar washback som de effekter proven har på hur inlärarna anvisas förbereda sig för ett prov. Washback kan också betyda den information som studenterna får efter provet om sina egna svagheter och starka sidor. Enligt Brown borde ett prov medföra washback och på så sätt fungera som ett verktyg för inläring. Inkorrekt svar borde visa vad som måste övas i framtiden och korrekta svar borde berömmas. Ett sätt att öka washbackeffekten är att kommentera provutförandet detaljerat och inte bara skriva betyget på pappret. Det är viktigt att ge feedback både på

deltagarnas starka sidor och på svagheter. Studenterna borde också ges möjlighet att kunna diskutera evalueringen med evaluaren. (ibid. 29)

Bachman & Palmer (1996) tar också upp testeffekten av språkprov. Testeffekt betyder de olika sätten ett prov påverkar samhället, utbildningssystemet och individer inom dessa. Den sista egenskapen som definierar ett provs genomförbarhet är hur praktiskt det är. Det betyder hur ett prov förverkligas i en viss situation och relationen mellan de resurser som krävs vid planering av ett test samt vid användningen av provet. Resurserna kan till exempel vara mänskliga eller materiella. Tid är också en resurs. Ett praktiskt prov kan inte kräva mera än de tillgängliga resurserna. (Bachman & Palmer 1996: 40.)

## **4.8 Den europeiska referensramens syn på bedömning**

Den europeiska referensramen (Eurooppalainen viitekehys 2003: 242) nämner tre viktiga saker om bedömningen, nämligen validitet, reliabilitet och användbarhet. Med validitet avses i referensramen att ett prov mäter det som det borde mäta i dess kontext. Ett prov är valitt om tolkningar av provresultaten stöds av teorin och empirisk evidens av företeelsen i fråga. Det betyder, att information som man får genom ett prov representerar de testades språkkunskaper.

Reliabilitet betyder enligt referensramen (Eurooppalainen viitekehys 2003) att ett prov är pålitligt när resultaten är likadana om man skulle använda provet på nytt med samma personer eller om andra personer skulle fungera som bedömare. Man måste också använda rätt standard när man mäter språkförmåga för att kunna få ett valitt resultat. Till exempel måste man kunna välja en viss CEFR-nivå (som B1) som standard vid bedömningen i en viss kontext. I den europeiska referensramen (ibid.) definieras 13 olika typer av bedömning (se tabell 5 nedan). Man kan till exempel värdera inlärningsresultat: det som man har lärt sig t.ex. under en viss kurs. Vid bedömning av språkkunskapsnivån värderas vad någon kan göra när det som han lärt sig tillämpas i det verkliga livet. CEFR-skolorna kopplar ihop bedömningen av språkkunskapsnivån och förmågan att uppfylla kraven i det verkliga livet. De tretton bedömningssätten framgår av tabellen nedan.



**Tabell 6: Bedömningssätt**

1	Bedömning av måluppfyllelse	Bedömning av språkfärdighet
2	Normrelaterad bedömning	Kriterierelaterad bedömning
3	Kriterierelaterad bedömning godkänt/ ej godkänt	Kriterierelaterad bedömning kontinuum
4	Kontinuerlig bedömning	Bedömning vid fasta tidpunkter
5	Formativ bedömning	Summativ bedömning
6	Direkt bedömning	Indirekt bedömning
7	Färdighetsbedömning	Kunskapsbedömning
8	Subjektiv bedömning	Objektiv bedömning
9	Bedömning enligt en checklista	Bedömning på en skala
10	Intuitiv bedömning	Styrd bedömning
11	Holistisk bedömning	Analytisk bedömning
12	Seriebedömning	Kategoribedömning
13	Andras bedömning	Självbedömning

Europeisk referensram för språk (2007: 181)

I tabell 6 ovan finns tretton bedömningssätt. Det har ingen betydelse på vilken sida begreppen har placerats. Varje sätt att se på bedömningen har två sidor i tabellen. Det mest relevanta synsättet på bedömningen för mitt arbete är nummer 2 i tabellen.

Enligt referensramen (Eurooppalainen viitekehys 2003: 251) kan bedömningen vara normrelaterad eller kriterierelaterad. Normrelaterad bedömning betyder att inlärnarna rangordnas och bedöms i relation till en jämförelsegrupp. Normrelaterad bedömning används till exempel i de finska studentskrivningarna. Kriterierelaterad bedömning är en reaktion mot normrelaterad bedömning. Då evalueras inläraren enligt den kapacitet/begåvning som han eller hon har visat i läroämnet, oberoende av jämförelsegruppens prestationer. Skalorna i referensramen och i GFL 2004 är kriterierelaterade.

Bedömningen kan vara kontinuerlig eller ske vid en fast tidpunkt (bedömningssättet nummer 4 i tabell 6). Bedömningen kan vara subjektiv eller objektiv (nummer 8). Subjektiv bedömning baserar sig på bedömarens omdöme. Vanligen betyder det evaluering av prestationens språkliga kvalitet. Davies et al. (1999: 191) förklarar subjektiv poängsättning som en procedur som innehåller bedömarens evaluering där alla möjliga svar inte är specificerade. Subjektiva testdelar kan variera mellan korta svar, där bedömarens omdöme är begränsat, och uppgifter som innehåller komplext utförande där den potentiella oenigheten mellan bedömare är stor. I objektiv bedömning finns det enligt referensramen (2003) bara ett korrekt svar, som till exempel i flervalsuppgifter.

Man kan bedöma direkt eller indirekt. Direkt bedömning betyder att man värderar det som den testade gör precis då, vid den tidpunkten. Till exempel intervjuer är direkt bedömning, medan ett lucktest är en form av indirekt bedömning. Man kan göra en skillnad mellan bedömningen av prestationsförmågan och kunskapsförmågan. Prestationsförmåga kan man evaluera med performanstester, vilka mäter den praktiska kompetensen. Uppgifterna i HY-TALK mäter prestationsförmåga och bedömningen är subjektiv och indirekt.

#### **4.9 Språkkunskapernas delkomponenter i en analytisk bedömning**

Till språkkunskaperna hör massor av olika kompetenser. I HY-TALK-bedömningarna gjorde bedömarna en analytisk bedömning för muntlig kompetens som innehöll skilda delbedömningar för genomförande av uppgiften, flyt, uttal, bredd, korrekthet och interaktion. Sist gjorde bedömarna också en holistisk bedömning av elevprestationen. I intervjuerna fick jag ett intryck av en viss polarisering mellan å ena sidan interaktion och flyt, och å andra sidan grammatiken. Få av mina informanter sade att de betonade grammatiska färdigheter i bedömningen. Jag skall i det följande ta upp ett specialproblem gällande korrekthet och grammatik, eftersom en av mina informanter tog upp korrekthetens betydelse vid bedömning av muntliga språkfärdigheter. Jag behandlar interaktion och kommunikation samt flyt ingående, men vill ta med i analysen också andra språkliga drag som bedömdes.

Talat och skrivet språk skiljer sig betydligt från varandra. Luoma (2004: 10) tar upp pauser, ton, omfång och intonation samt variation i dem som exempel på hur talaren kan uttrycka mening och nyanser. Den europeiska referensramen beskriver fonologin och uttalet som följande. Fonologisk kompetens (2007: 113) innehåller enligt gemensam europeisk referensram kunskaper och insikt om samt produktion av:

- Fonemen i språket i fråga samt deras realisation som varianter i vissa kontexter som allofoner.

- Fonetiska särdrag som skiljer fonemen från varandra, till exempel sonoritet, rundning och nasalitet.
- Den fonetiska uppbyggnaden av ord, som innehåller stavelsestruktur, fonemsekvens, betoning och ton.
- Prosodi, som uppfattar meningens och satsens betoning, rytm och intonation.
- Fonetisk reduktion som innehåller vokalreduktion, starka och svaga former, assimilation och elision (utelämning).

I tabell 7 nedan finns beskrivningen av fonologisk kontroll från CEFR. På A-nivåerna har uttalet en klar utländsk accent och uttalet kan ännu vara otydligt. På B1-nivån är uttalet ”tydligt förståeligt”, men en utländsk accent är ibland synlig och det förekommer fel i uttalet. På B2-nivån borde uttalet och intonationen redan vara naturlig och tydlig. På C-nivån behärskar språkinläraren satsbetoningen och kan uttrycka finare nyanser i betydelsen.

**Tabell 7: Fonologisk kontroll, beskrivning av fonologisk kompetens enligt Europeisk referensram för språk (2007: 113)**

	BEHÄRSKNING AV UTTAL OCH INTONATION
<b>C2</b>	<i>Som C1</i>
<b>C1</b>	<i>Kan variera intonation och använda korrekta betoningar i meningen för att uttrycka fi nare betydelsenysanser</i>
<b>B2</b>	<i>Har tillägnat sig ett tydligt och naturligt uttal och intonation</i>
<b>B1</b>	<i>Uttalet är helt begripligt även om det ibland hörs en brytning och enstaka felaktiga uttal förekommer.</i>
<b>A2</b>	<i>Uttalet är i stort sett tillräckligt tydligt för att man ska förstå, även om det fi nns en tydlig brytning och samtalspartnern ibland kan behöva be om upprepning.</i>
<b>A1</b>	<i>Uttal av ett mycket begränsat förråd av inlärd ord och fraser kan förstås med viss ansträngning av modersmålstalare som är vana att träffa personer från hans/hennes språkområde.</i>

I CEFR (2001: 128) definieras flyt (*fluency*) som en del av den funktionella förmågan i vilken också precision i påståenden (kan formulera påståenden så att man gör sin mening

tydlig) ingår. Flyt betyder förmåga att artikulera, gå vidare i interaktion och klara av situationer när man hamnar i en återvändsgränd (*dead end*).

**Tabell 8: Muntlig flyt enligt Europeisk referensram för språk (2007)**

	<b>FLYTANDE FRAMSTÄLLNING (MUNTIG)</b>
<b>C2</b>	<i>Kan uttrycka sig utförligt med ett naturligt och ledigt fl yt utan att tveka. Pausar endast för att hitta exakt rätt ord för att uttrycka sina tankar eller finna ett lämpligt exempel eller förklaring.</i>
<b>C1</b>	<i>Kan uttrycka sig ledigt och spontant, praktiskt taget utan ansträngning. Endast begreppsmässigt komplicerade ämnesområden kan förhindra ett naturligt och flytande språk.</i>
<b>B2</b>	<i>Kan kommunicera spontant och då ofta uttrycka sig anmärkningsvärt flytande och ledigt, även i längre, komplicerat och sammanhängande tal.</i>  <i>Kan tala sammanhängande i ett ganska jämnt tempo. Trots att han/hon ibland kan tveka när han/hon letar efter satsmönster och uttryck, förekommer få märkbara längre pauser.</i> <i>Kan fungera i samtal så pass flytande och spontant att ett normalt umgänge med modersmålstalare blir helt möjligt utan ansträngning för någondera parten.</i>
<b>B1</b>	<i>Kan uttrycka sig relativt lätt. Trots vissa formuleringsproblem som resulterar i pauser och återvändsgränder kan han/hon fortsätta utan hjälp.</i>  <i>Kan tala sammanhängande och göra sig förstådd, även om det är uppenbart att han/hon ibland stannar upp för att planera nästa grammatiska och lexikala steg eller rätta till fel, särskilt vid längre avsnitt av fritt tal.</i>
<b>A2</b>	<i>Kan göra sig förstådd i korta inlägg, även om pauser, omtagningar och omformuleringar är mycket markanta.</i>  <i>Kan konstruera fraser om kända ämnen tillräckligt lätt för att klara av korta samtal, även om tvekan och omtagningar är mycket tydliga.</i>
<b>A1</b>	<i>Kan använda mycket korta, fristående, främst inlärda uttryck. Stannar ofta upp för att leta efter uttryck eller uttala mindre kända ord och för att rätta till missförstånd.</i>

I tabell 8 ovan nämns exempel på flyt och inlärares problem när det gäller flyt. På A-nivån finns det tveksamhet, pauser, fel början och reparationer. Flyt finns på frasnivå och uttrycken är korta. På B-nivån nämns pauser speciellt i längre produktion av fritt tal, men inlärares kan uttrycka sig relativt lätt och kan hålla taktakten med vissa avvikelser. På C-nivån kan inlärares uttrycka sig flytande och spontant utan större ansträngning.

## 5 Resultat

I detta avsnitt skall jag presentera resultaten av intervjuerna samt av lärarnas elevbedömning och svara på mina forskningsfrågor. I avsnitt 5.1 kommer jag att ge svar på frågan om hur två grupper av bedömare upplever bedömningen av elevernas muntliga språkkunskaper i svenska enligt nivåskalan för språkfärdigheter. I avsnitt 5.2 skall jag svara på den andra forskningsfrågan: hurdan är reliabiliteten mellan bedömarna i bedömningen enligt nivåskalan?

### 5.1 Bedömarna i fokus

Alla informanter jobbar på universitetsnivå och har använt nivåskalan för språkfärdigheter eller en annan finsk CEFR-baserad skala i sitt arbete tidigare, förutom en informant som för första gången i projektet HY-TALK använde nivåskalan. Informanterna har dock för det mesta använt olika nivåer i skalorna än vad man använde i HY-TALK- evalueringen. Alla lärare förutom en har ändå använt skalorna tidigare i sitt arbete och vet hur de fungerar. De är alla professionella språkbedömare. Alla informanter förutom en har också haft utbildning i att använda CEFR-skalorna, men den har mest gällt skriftlig bedömning.

Det fanns också skillnader mellan lärarnas erfarenheter i CEFR-bedömning. Jag litar på lärarnas egen deskription av sin erfarenhet. Erfarenheten är visserligen subjektiv och innehåller informantens egen subjektiva uppfattning om sin erfarenhet i bedömningen enligt CEFR-baserade skalor. Det är dock omöjligt att objektivt mäta och jämföra informanternas erfarenhet i CEFR-bedömning, eftersom informanterna har olika bakgrunder och har använt olika nivåer i nivåskalan i sitt bedömningsarbete. Dessutom är jag intresserad av att studera om det finns skillnader i elevbedömningen mellan informanter som förhåller sig olika till användningen av nivåskalan i bedömningen. Jag vill alltså jämföra om de bedömare som känner sig som erfarna bedömare bedömer olika än de som känner sig oerfarna. Eftersom mitt material baserar sig på två olika bedömningssessioner, kan jag bara jämföra resultaten inom de sessionerna. Detta är en

brist men en praktisk sak som jag inte kan påverka. Jag skall också ta upp lärarnas kommentar om reliabilitet och validitet i resultaten.

Jag har valt att beskriva intervjuresultaten i form av bedömarprofiler på basis av lärarnas egna berättelser och betoningar i intervjusituationen. Lärare 1 betonade, att hon inte har mycket erfarenhet av att bedöma muntliga språkfärdigheter med hjälp av CEFR-skalar. Lärarna 2 och 3 sade att de är erfarna bedömare, men har för det mesta använt olika nivåer i skalan än de som används i HY-TALK-sessionen. Lärare 4 sade också att hon har erfarenhet av att bedöma med nivåskalan. Lärare 5 litade på sig själv som bedömare men sade att hon vanligtvis inte använder skalan så noggrant som i HY-TALK. Lärare 6 hade inte använt skalan tidigare, och kände sig inte alls erfaren, men förhöll sig ändå positivt till skalan. Lärare 7 berättade att hon hade mycket erfarenhet i att bedöma enligt en CEFR-baserad skala, och hon hade till och med varit med i att utveckla nationella skalar på högskolenivå.

Bedömarna 1-4 deltog i bedömningssession 1 i vilken de bedömde elever på gymnasienivå. Bedömarna 5-7 tog del endast i bedömningssession 2 där de bedömde studerande på högskolenivå.

**Tabell 9: Bedömarprofil, bedömare 1.**

Erfarenhet som bedömare	Bedömare 1 säger att hon har väldigt lite erfarenhet av CEFR-bedömning. Hon har använt det i HY-Talk-projektet och fått utbildning i att använda det. Hon har bedömt skriftliga prestationer men inte muntliga.
Reliabilitet	Hon nämner att i kriterierna för interaktion talas det om ”basförråd av strategier”. Man kan tolka det på olika sätt när man evaluerar. Enligt bedömare 1 är elevens personlighet med i bedömningen. Hur rädda eleverna är inför kameran och hur vana de är vid en sådan här situation spelar säkert en viss roll.
Validitet	Bedömaren kunde inte mäta kunskaper med skalan så noggrant som hon skulle ha velat. Det var inte alltid möjligt att med hjälp av skalan få fram en skillnad som man tyckte att fanns mellan olika studenter eftersom skalan är så allmän.  Bedömaren problematiserade bedömningen av muntlig färdighet. Betyder det bara

	<p>att man ska bedöma produktion, korrekthet och lexikon, eller ska man ta dialog med i provet?</p> <p>Enligt Bedömare 1 är det hur man kommunicerar allra viktigast i bedömning av muntlig färdighet. Då måste man beakta flera parametrar hela tiden i bedömningen.</p> <p>Bedömare 1 kritiserade interaktionskriterierna. Enligt henne kunde eleverna ”utnyttja basstrategier för att hålla diskussionen i gång”. Eleverna var alltså till och med på nivå B2 i interaktion enligt kriterierna. Kriterierna undervärderar elevens interaktionsförmåga.</p> <p>Olika slags personligheter får olika bedömning enligt Bedömare 1. Pratglada personer klarar sig bättre. Personligheten påverkar bedömningen eftersom hela personen är med i situationen.</p>
Autenticitet	<p>Noggranna instruktioner gör uppgifterna mindre lika autentiska situationer enligt bedömare 1. Om man måste följa strikta instruktioner, mäter provet förmågan att översätta vissa uttryck till svenskan. I en autentisk situation skulle man använda olika strategier för att hålla diskussionen igång, t.ex. kringgå ordet, förenkla, fråga osv. I det verkliga livet finns det inte alltid manus för diskussionen. Bara den sista uppgiften (på gymnasienivå) liknade sådant språkbruk som man sysslar med utanför skolan.</p>
Genomförande av uppgiften	<p>Bedömaren tycker att uppgifterna med inga färdiga repliker är bättre än sådana som har strikta instruktioner. Då är det svårt för eleven att kunna visa att man är över den nivån som krävs i uppgiften. ”Att komma överens om en utflyktsdag” var på så sätt bra att den var bättre formulerad än uppgift 2. En monologisk uppgift var enklast eftersom man inte behövde ta hänsyn till de andra utan kunde bara tala till kameran. Bedömaren betonade interaktionen i uppgifterna.</p>
Flyt	<p>En del av eleverna fastnade bara i att översätta replikerna till svenska. Då blev det väldigt långa pauser och hackigt, vilket påverkade bedömningen negativt.</p> <p>Eleverna beaktade instruktionerna för noga och följde dem ordagrant. Detta gjorde flytet sämre och gjorde att nivån blev lägre.</p>
Uttal	<p>Bedömaren önskar att det fanns språkspecifika och mera noggranna skalor. Hon tycker att det borde definieras att vad som avses med flyt och uttal i svenskan på de olika nivåerna. Hon nämner som exempel u-ljudet: på vilken nivå ska man kunna u-ljudet, eller en bra frاسبetonning.</p>

Bredd	På lägre nivåerna i skalan har korrekthet och bredd större betydelse. Då börjar felen hoppa fram. Från nivå A2 uppåt betonade bedömaren helhetsgenomförande av uppgiften och interaktionen i sin bedömning.
Grammatisk korrekthet	<p>Böjningsfelen har ingen stor betydelse om man gör sig förstådd, tycker bedömaren 1.</p> <p>Bedömaren önskar att det fanns språkspecifika och mera noggranna skalor. Hon tar som exempel kriterierna på nivå A2.1 ”virheitä esiintyy yhä paljon perusrakenteissakin”. Vad är de grundläggande strukturerna man ska kunna? Starka verbens böjning (kom / komde)? På vilken nivå kommer ordföljden i huvudsatsen? På vilken nivå skall man kunna producera större fraser?</p> <p>Bedömaren 1 tycker att skalorna är för allmänna och grammatiska strukturer och regler borde sättas på skalans olika nivåer. Hon vill ha en nivåskala speciellt för svenska språket, där exempel på de olika grammatiska strukturerna skulle finnas och de skulle ha kategoriserats enligt skalans nivåer.</p> <p>Bedömaren är inte säker på var de olika grammatiska strukturerna hör i skalan. Hon vet inte vad man borde kunna på de olika nivåerna.</p>
Interaktion/Kommunikation	<p>Interaktionen påverkar alla delfaktorerna i bedömningen (bredd, korrekthet osv.) enligt bedömarens åsikt. Beskrivningen av interaktionen i skalorna undervärderar inlärares interaktionsförmåga. Eleverna kan föra ett samtal på svenska fast de kunde inte tala det så bra. De kanske inte gör det på ett helt korrekt sätt, men det är inte viktigt för bedömaren.</p> <p>Om interaktionen inte tas med, ger det en väldigt förlegad syn av vad muntlig färdighet är. Enligt bedömaren 1 är det viktigt i muntliga prov att beakta det hur man kan agera med en annan människa. ”Det är det första som frågas sen utanför skolan, om man kan använda språket i en verklig situation med någon annan människa, och inte bara så att man står och rabblar upp en massa ord.”</p>
Speciellt för denna bedömaren	<p>Bedömaren 1 önskar mer detaljerade och språkspecifika kriterier.</p> <p>Instruktionerna av uppgifter borde vara lösare (inga färdiga repliker) i HY-Talk för gymnasieelever enligt henne, och kanske till och med på svenska.</p>

**Bedömaren 1** har ett interaktivt och socialt synsätt på språket (se tabell 9). Hon betonar interaktion och kommunikationsförmåga i provet. Hon ifrågasatte skalornas validitet. Hon kritiserar styrda uppgifter, och enligt min tolkning tycker hon att friare uppgifter



vore bättre och mera autentiska och på det sättet skulle göra provet mera valitt. Hon önskar också att skalorna skulle anpassas och utvecklas till språkspecifika skalor med mera konkreta exempel av t.ex. grammatiska strukturer på olika nivåer. Hon betonar den pragmatiska kompetensen (se figur 2 på sida 28) i intervjun.

**Tabell 10: Bedömarprofil, bedömare 2**

Erfarenhet som bedömare	Bedömare 2 har använt skalan på universitetsnivå, har fått utbildning i att använda skalan (av studentexamensnämnden).
Reliabilitet	<p>Man bedömer först ganska olika jämfört med andra bedömare i en bedömningssituation, tycker bedömare 2. Bedömaren sade att hon behöver en tid att zoma in och få syn på att vad det är hon ska bedöma. En mera långsiktig bedömning är lättare om man känner sina studenter från tidigare.</p> <p>Hon säger att bedömningen av muntliga språkfärdigheter aldrig blir riktigt rättvist.</p> <p>Hon problematiserade bedömarrollen moraliskt och etiskt: ”vad anser mitt samvete att den här människan ska få som resultat”. Bedömningssituationen är obehaglig och svår för henne eftersom hon måste värdera någon annan människa och ge ett vitsord åt henne.</p> <p>Att ha flera som bedömer en elev gör bedömningen lättare och lite mera rättvis. Olika bedömare prioriterar olika beroende på hurdan personlighet man är. Man värderar olika saker i bedömningen.</p> <p>Hon tycks ha svårigheter att bestämma vad det är man skall bedöma: språkkunskaper eller nervositet eller hela människan.</p>
Validitet	<p>Kriterierna och provsituationen mäter också någonting annat än kommunikation och muntlig färdighet, till exempel vem som är nervös i en stressad, onaturlig situation, säger bedömare 2. Situationen är mera att uppträda, vilket inte beaktas av kriterierna. Provet mäter inte det som det ska mäta. Bedömare 2 ifrågasatte skalornas användning i en oautentisk provsituation.</p> <p>Bedömare 2 konstaterar nog också att skalorna tar upp olika nyanser i muntligt språkbruk och hjälper bedömaren se på olika saker i kommunikationen och inte till exempel bara på korrekthet. Det är också viktigt enligt henne att diskutera</p>

	bedömningen med en annan bedömare för att göra bedömningen bättre.
Autenticitet	Bedömaren problematiserade hela provsituationen. Enligt henne mäter den inte bara språkkunskaper utan också avslappning inför kameran osv. För henne var det problematiskt att provsituationen inte är en autentisk situation, men att man bedömer prestationen som om det vore det.
Flyt	Bedömare 2 problematiserade också begreppet flyt. Det kan variera mellan olika människor enligt henne, du kan ha en viss rytm i ditt språk, och rytmen är olika.
Uttal	Vad anses vara störande uttala och enligt vem då? Bedömare 2 ifrågasatt krav för uttalets korrekthet. Hon tycker också att bedöma uttalet och till exempel flyt samtidigt är krävande eftersom man måste beakta många olika saker samtidigt.
Grammatisk korrekthet	För Bedömare 2 är korrekthet inte så väsentligt när det gäller muntlig färdighet, men om fel i korrekthet förorsakar brister i kommunikationen, då är det ett fel. Grammatiska fel är alltså väsentliga om man inte lyckas få fram sitt budskap.
Interaktion/ Kommunikation	<p>Elevernas interaktiva beteende påverkade Bedömare 2:s bedömning. Kommunikationen är det viktigaste när man bedömer muntlig färdighet.</p> <p>Skalorna utgår från att de som samtalar är på samma nivå, men enligt Bedömare 2 är situationen en annan om den ena har språket som modersmål och den andra inte. ”Då är situationen också en annan än om jag till exempel som har svenska som modersmål så talar med dig för att då kan jag fylla i sådant som du säger och där igen har kommunikationen en annan dimension.” På grund av en paruppgift kan man inte veta om paret klarar sig i en situation med modersmålstalare.</p> <p>Enligt henne är det också svårt att dra gränsen mellan språk och icke-språk: vad ska man bedöma och vad inte. Det är svårt att bedöma kommunikation.</p>
Speciellt för denna bedömare	<p>Föreslår en spalt för ”grad av nervositet” för att göra den onaturliga provsituationen synlig i bedömningen.</p> <p>Skillnaderna mellan olika nivåer i skalan är inte tydliga enligt henne. ”jag tycker att man ibland skärper det på ena nivån och nästan löser det på den andra nivån”</p>

**Bedömare 2** har också en kommunikativ och sociolingvistisk syn på språkkunskaper. Hon betonar interaktionen i sin bedömning men kritiserar hårt provsituationens

onaturliga karaktär. Hon tycker att kriterierna beskriver autentiskt språkbruk och därför kan de inte användas utan problem i en icke-autentisk provsituation. Nervositet och avslappningsgraden kan variera och påverka prestationen. Autenticitet och provets validitet är problematiska för henne, men hon talar inte om mera eller mindre autentiska situationer, utan om autentiska och icke-autentiska. Nedan finns ett citat ur hennes intervju. Enligt henne finns det inte fel i kriterierna i skalan i sig, men de används på fel plats. En provsituation är inte en autentisk situation. Provet mäter inte det som det ska mäta, utan också någonting annat. Hon nämnde nervositet som exempel. Kriterierna beaktar inte att det är en onaturlig situation för att kommunicera, det är mera att uppträda. Kriterierna är inte synkroniserade enligt henne.

*”När folk sitter och tittar på dig när du ska prata, så har det inte riktigt med kommunikation att göra, det är mera att uppträda. Det är en annan genre och det beaktas inte av kriterierna. ... Jag tror inte riktigt på det helt och hållet, att det mäter det det ska mäta.”* (Bedömare 2.)

Hon talar också indirekt om reliabilitet: hon säger att det finns variation mellan bedömare: bedömare lägger märke till olika saker i bedömningen. För henne tycks det inte vara klart om ska man bedöma hela människan, dennas performans och nervositetsgrad, eller språkkunskaper. Hon problematiserar bedömningen av den pragmatiska kompetensen.

**Bedömare 3** har en kommunikativ syn på bedömning av muntliga språkkunskaper. Hon betonar pragmatisk kompetens i intervjun och föreslår att nonverbal kommunikation tillsätts som ett eget drag i den analytiska bedömningen inom HY-TALK. Kommunikation och dess naturliga drag är viktiga för henne i muntlig språkfärdighet och hon tar kommunikationen med i bedömningen.

**Tabell 11: Bedömarprofil, bedömare 3**

Erfarenhet som	Har tidigare använt skalorna i sitt arbete som bedömare av muntlig färdighet. Har dock tidigare använt olika nivåer än i HY-TALK, och tyckte att det var därför
----------------	---

bedömare	jobbigt att bedöma.
Reliabilitet	<p>Bedömaren tyckte att hon var för snäll i början och blev strängare mot slutet.</p> <p>Bedömare 3 tyckte att det fanns jättelite spridning mellan eleverna, nästan alla var på samma nivå.</p> <p>Bedömaren önskar att man kunde se på elevprestationerna flera gånger innan man bedömer. Hon tycker att efter varje par blev det lättare att bedöma. Man måste lyssna på många drag simultant, vilket kräver mycket koncentration, och risken finns att man missar någonting. Hon tror ändå inte att hon skulle ha bedömt olika efter att ha lyssnat på elever tre eller fyra gånger.</p> <p>Bedömaren säger, att hon vet från praktiken att vissa lärare är snällare och andra strängare. ”Man kan se på samma kriterier och någon tolkar dem strängare och någon annan snällare”. Det kräver långvarigt arbete för att alla ska komma på samma nivå.</p> <p>”Ju mera erfaren man är att använda skalan, desto säkrare är man i bedömningen.”</p> <p>Bedömare 3 tycker att det är viktigt att föra diskussion med kolleger om bedömningen. Om man har till exempel ett par där den ena talar mycket och visar bra kommunikativa kunskaper, medan den andra inte säger mycket men talar grammatiskt korrekt, kan man diskutera om bedömningen och besluta med sina kollegor vem som ska få högre vitsord. En bedömare kan vara mera känslig för någonting och en annan för någonting annat i prestationen. Bedömare 3 tycker också att det finns risk att bedömaren är ouppmärksam en stund i bedömningssituationen och då kan t.ex. ett misstag gå förbi i elevprestationen.</p> <p>Bedömaren undrade om hon hade använt för smal skala. Å andra sidan säger hon, att om eleverna är på en jämn nivå, har läraren gjort ett jämt och bra jobb och att skolundervisningen i så fall har lyckats. Enligt henne använder en erfaren bedömare bredare skala än en oerfaren. Hon tycker att CEFR-skalan gör bedömningen jämnare och inte subjektiv, när man har objektiva kriterier som man följer. ”Om vi kan använda skalan rätt så hjälper det nog verkligen att ge ett rättvisare vitsord än tidigare.”</p>
Validitet	<p>Bedömaren lyssnar på uttalet och fonetik i sin bedömning. Hon nämner också interaktionen och bredden av ordförråd och konstruktioner. Hon konstaterar att bedömningen ändå är en intuitiv helhetsbedömning, och inte matematisk uträknat från delbedömningarna.</p> <p>Bedömare 3 tycker, att en detaljerad och noggrann bedömning som man gjorde i</p>

	<p>HY-TALK är ett bra sätt att gå vidare, och att hon tycker om det.</p> <p>Bedömaren tycker att det är problematiskt att när man kommer uppåt på skalan så krävs det mera av varje drag jämnt. I verkligheten kan man ändå ha stort ordförråd men dåligt uttal till exempel. Om man måste ge ett enda vitsord så måste man väga vilka som är de viktigaste dragen. Bedömare 3 tycker att den kommunikativa aspekten är viktig. Hon föreslår en spalt till skalorna som skulle innehålla nonverbal kommunikation.</p>
Genomförande av uppgiften	<p>Uppgifterna var styrda och därför var det lite som eleverna kunde variera i sin output, tycker bedömare 3. Därför var det svårt att veta om eleven var på nivå A2.1 eller A2.2, eftersom man fick så lite evidens på vad eleven kan. Uppgifterna krävde inte mycket kommunikation.</p> <p>Detaljerade anvisningar var ett problem i provsituationen. Provsituationen skulle må bättre av att ha någon del där de får vara lite friare.</p>
Uttal	För bedömaren är interaktionen, uttalet och fonetik det viktigaste i bedömningen.
Bredd	Bredden av ordförråd och konstruktioner är viktigt i bedömningen för bedömare 3.
Grammatisk korrekthet	<p>Bedömaren säger att hon inte vet vad begrepp korrekthet syftar på: på grammatiska konstruktioner eller någonting annat.</p> <p>För bedömaren är det inte viktigt om man gör grammatiska fel eller inte om man nonverbalt kan kommunicera så bra att mottagaren ändå förstår. I uppgiftssituationerna kan man klara sig mest med nonverbal kommunikation.</p> <p>Vissa strukturella fel (nämner som exempel fel pronomen som <i>er</i> och <i>vår</i>, <i>min</i> och <i>din</i>, <i>köpt</i> och <i>sålt</i>) stör kommunikationen och kan påverka helheten så att det blir kommunikativt fel. Till exempel: "<i>Hur mycket pengar ska vi ge</i>" när det är meningen att ta med pengar. Vissa fundamentala strukturella fel som t.ex. att blanda vissa pronomen och verb kan förorsaka kommunikativa fel och bli farliga.</p>
Interaktion/Kommunikation	Ju längre ner på skalan man finns, desto viktigare är den kommunikativa biten för Bedömare 3. "Det är inte så farligt om du säger fel eller rätt, bara om det blir någonting. När vi kommer på C-nivån så måste det också vara rätta ord.
Speciellt för denna bedömare	Det är svårt att göra skillnad mellan nivåer A2.1 och A2.2 angående flyt, lexikon och strukturer för bedömare 3. Samma problem när det gäller att tala på skalan (samma nivåer).

	<p>Bedömaren tycker att fel i enstaka ord i diskussionen kan förorsaka fundamentala fel i kommunikationssituationen. Hon talar inte om mera avancerade grammatiska fel. Hon tycker också, att vissa situationer kan man klara av också mest med nonverbal kommunikation. Hon föreslår att i bedömningen en spalt för nonverbal kommunikation skulle tillsättas. Nonverbal kommunikation stöder den verbala kommunikationen. Hon menar att man också skulle bedöma till exempel ögonkontakt och gester som ersättningar för ord. Hon nämner att några par hade ingen ögonkontakt och tittade bara på sina papper, vilket resulterade i dålig kommunikation.</p>
--	--

**Bedömare 4** betonade flyt, talhastighet och interaktion i intervjun. Hon ser en klar skillnad i bedömningsarbetet nu och tidigare före CEFR-skalorna. Då var bedömningen subjektivare enligt henne.

**Tabell 12: Bedömarprofil, bedömare 4**

Erfarenhet som bedömare	<p>Bedömare 4 har jobbat med CEFR-skalorna tidigare i sitt arbete och i studentexamensnämnden. Hon har bedömt muntliga elevprestationer tidigare med en CEFR-baserad skala. Hon har ändå mest jobbat med andra nivåer i skalan än i HY-TALK-evalueringen.</p>
Reliabilitet	<p>Bedömaren tyckte att hon var långsammare i bedömningen i HY-TALK-sessionen, men att bedömningen gick bra. Hon tycker att alla bedömare var ganska samstämmiga den gången, och inga stora skillnader fanns. Bedömare betonar ändå olika saker enligt henne. Hon hittade otydligheter i vissa formuleringar i skalorna.</p> <p>Bedömare 4 tycker att skalan underlättar diskussionen mellan bedömare, och under diskussionen kan man märka om man bedömer olik de andra. Hon säger också att man måste arbeta med bedömning för att kunna vara en bra bedömare och för att lägga märke till att andra lägger mera vikt på någonting än vad man själv gör.</p>
Validitet	<p>Förr var bedömningen ganska subjektiv, men skalorna har gett ett hjälpmedel till att man kan se mera objektivt på elevprestationen och beakta många olika aspekter i språkkompetensen, säger bedömare 4. Hon tycker att skalorna är ett verktyg som gör bedömningen mycket mera valid och mindre subjektiv. Som bedömare tänker man kanske inte hur många faktorer som det finns att man måste beakta. Med hjälp av skalorna är det enklare att motivera och ge respons till eleven. Skalorna har hjälpt bedömare 4 mycket. Man vill ge en holistisk bedömning men det är bra att göra en analytisk bedömning också. Diskussionen med kolleger blir mera analytisk, man har begrepp att tala om saker. Skalan minskar också jämförelse mellan elever. Elever kan ha olika styrkor och profiler men skalorna gör det lättare</p>

	att bedöma vad just den personen kan och inte jämföra med andra.
Genomföra nde av uppgiften	Bedömaren tycker, att styrda uppgifter i HY-TALK inte gav möjlighet för eleven att komma på en högre nivå än det som krävs i uppgiften. En friare uppgift är viktig enligt henne, eftersom då kan man visa mera av sig. Hon ger som exempel en elev som lät flytande men använde samma ordförråd hela tiden, tex ordet <i>jätte</i> . I en styrd uppgift blev det otydligt om hon skulle ha kunnat mera eller om hon hade begränsat ordförråd.
Flyt	<p>Bedömaren fäster uppmärksamhet vid att eleven låter svensk, att det finns ett visst flyt. Bedömaren tycker att hon betonar flyt när hon bedömer muntlig färdighet. Hon nämner också talhastighet och att man kan använda undvikande strategier och undvika pauser som viktiga element i hennes bedömning.</p> <p>Bedömaren tycker att skalorna för uttal är krävande och grova. Redan på en låg nivå borde man kunna väldigt bra. Hon tycker att det är också svårt att definiera vad flyt är. ”Man måste ha ganska mycket bakgrundskunskap för att veta hur bedöma när det gäller uttal eller flyt.”</p>
Uttal	Bedömaren tycker att det är problematiskt att bedöma uttal eftersom svenska kan ju låta ganska annorlunda i vissa dialekter. Vad ska man tolerera och vad är fel?
Bredd	Bedömaren 4 definierar bredd som ” att man kan använda varierande konstruktioner och har ett stort ordförråd ”. I HY-TALK är uppgiften så styrd att man inte kan visa bredden. Om någon är riktigt bra så har han svårt att visa det genom väldigt styrda uppgifter.
Grammatisk korrekthet	På lägre nivåer (A och B) är grammatiska saker inte så viktiga för bedömaren. På C1-nivån är det viktigare enligt Bedömaren 4 att kunna uttrycka sig rätt och ha rätta grammatiska konstruktioner. På begynnande nivån är det viktigaste enligt henne att man klarar av situationen. På C-nivån blir mindre nyanser och idiomatiska uttryck viktigare. Man måste kunna använda dem rätt.
Interaktion/ Kommunikation	Enligt Bedömaren 4 är interaktionen väldigt viktig i provsituationen. Hon nämner som exempel ett par i HY-TALK som kunde hjälpa varandra i uppgifterna. Det tycker hon att är en viktig del. Interaktion är inte ett kriterium vid sidan av uttal och grammatik, utan genomsyrar allt. Det är svårt att bedöma andra faktorer utan att ta till hänsyn interaktion.
Speciellt för denna bedömaren	<p>Bedömaren förhåller sig speciellt positivt mot skalorna.</p> <p>På forskarnivå och i lärarutbildningen är skalan känd och man har kommit ganska långt med arbetet, men ute på fältet bland lärarna kan skalan kännas för tung att</p>

	<p>lära sig. Boken är inte lätt att läsa och den innehåller mycket information. Det går inte snabbt att tillägna sig skalorna, och ordentlig utbildning krävs för att lära sig använda dem och bedöma enligt dem. Utbildningsstyrelsen har ordnat kurser men problemet är att bara de mest aktiva lärare deltar i fortbildningen. Bedömare 4 tycker att lärarutbildningen spelar en viktig roll när det gäller lärarstuderande.</p> <p>På A- och B-nivåerna är det lätt att bedöma enligt skalan eftersom det finns mycket text i skalan. När man kommer på C-nivån, borde det vara mera text runt skalorna för att kunna lättare förstå vad skalorna innebär.</p>
--	--

Bedömare 4 nämner flyt, talhastighet och undvikande strategier som exempeldrag som hon lyssnar på i bedömningen av muntliga språkfärdigheter. Hon nämner både drag i språkkompetensen som hör till den organisatoriska kompetensen som uttal, och till den pragmatiska kompetensen, som till exempel undvikande strategier. Flyt kan också ses som en del av kontroll.

**Bedömare 5** har en mera kognitiv syn på språkkunskapernas bedömning i intervjusituationen. Hon betonar de kunskaper som är viktiga för elever i arbetslivet. Hon deltog i bedömningssession 2 och bedömde högskolestuderande.

**Tabell 13: Bedömarprofil, bedömare 5**

Erfarenhet som bedömare	Bedömare 5 har fått utbildning i att använda skalorna och använt dem i sitt arbete i ungefär fem år. Hon har också varit med i att utarbeta egna beskrivningar på sin arbetsplats.
Reliabilitet	<p>Bedömare tycker att bedömningarna i HY-TALK var ganska liknande, men hon tror att hon själv har varit partisk när det gäller hennes egna studenter som var med i HY-TALK-inspelningarna. Hon säger att sina egna studenter förstår man mycket bättre än andra. Hon baserar sin bedömning ändå på känsla och på erfarenhet, och tittar sällan på kriterierna. ”Jag tänker på hur det skulle fungera i praktiken, i en riktig situation i arbetslivet.”</p> <p>Bedömare 5 säger att kriterierna har påverkat hennes arbete så att de har dragit gränserna neråt och varit strängare tidigare. Skalorna kräver inte att man skall tala perfekt för att kunna få godkänt.</p> <p>Skalorna är positiva eftersom de gäller alla. Bedömaren tycker att vissa personer är strängare och andra snällare, men skalorna har vissa kriterier som alla i hela</p>



	Finland borde följa. Hon nämner ändå som ett problem att skalorna kan tolkas på många olika sätt. Det finns också skillnader mellan yrkeshögskolor och universitet när det gäller användning av skalor. Bedömaren tycker att olika slags tolkningar inte är illa om studenter också möter olika krav i verkligheten efter provet.
Validitet	<p>Bedömaren 5 tycker att skalorna hjälper bedömningen eftersom de innehåller exakta beskrivningar av kunskaper.</p> <p>Bedömaren tycker att analytisk bedömning av bredd, uttal osv. separat är bra eftersom det hjälper att specificera. Alla spalter var ändå inte relevanta i alla uppgifter. Hon ifrågasätter delen ”genomförande av uppgiften”. Hon tycker att hennes bedömning ändå blev intuitivt (”hatusta vedettyä”) och man borde ha tittat på prestationerna två gånger.</p> <p>Hon nämner att personlighet och kroppsspråket kan påverka bedömningen, vilket är negativt. Man skall utvärdera språkkunskaper, inte kroppsspråket eller personlighet. Det påverkar också att vad man hinner lägga märke till och på hurdant humör man är.</p> <p>Bedömaren säger många gånger under intervjun att hon har en osäker känsla ofta när hon bedömer, och vitsorden baserar sig på intuition. Då kan skalorna hjälpa att specificera.</p>
Autenticitet	Om personen är nervös kan hon prestera sämre än normalt. Man borde mäta kunskaperna flera gånger i olika situationer för att göra bedömningen mera rättvist.
Genomförande av uppgiften	Bedömaren tyckte inte att uppgifterna i HY-TALK var lyckade. Den infödda talaren i provsituationen var först inte bra förberedd, men blev van vid att diskutera och talade mera mot slutet, vilket var orättvist för paren. I första uppgiften fanns ingen interaktion fast bedömarna borde bedöma det. Texterna i uppgift tre passade bättre för vissa par (jurister) än andra. Hon säger också att många olika slags uppgifter i provet var en bra sak.
Bredd	Bedömaren nämner diskussion, ord och terminologi samt korrekthet som viktiga delar av en muntlig prestation.
Grammatisk korrekthet	Bedömaren säger att hon inte betonar grammatiska kunskaper hos studerandena fastän hon lyssnar på det också.
Interaktion/Kommunik	Personlighet och kroppsspråket får inte påverka bedömningen.

ation	
Speciellt för denna bedömare	<p>Bedömare 5 tycker att kraven har blivit lägre med skalan, man kräver inte så mycket som tidigare. Hon tycker att skalorna uppmuntrar studenter och har gett en positiv syn på språkkunskaper. Hon kollar kriterierna ganska sällan, bara en gång i terminen.</p> <p>Hon talar också om hur man bedömer sina egna studenter: om man känner den som bedöms från tidigare, kan man använda info utanför provsituationen när man bedömer. För henne är det inte bara provprestationen som definierar vitsordet. Bedömaren talar inte om interaktion eller skilt om de olika delbedömningarna, men nämner korrekthet, grammatik och ordförråd som viktiga saker i bedömningen.</p>

Bedömare 5 har ett mera kognitivt synsätt på språkinläring. Hon nämner inte interaktion i intervjun, och talar inte heller om kommunikativ förmåga eller kompetens. Hon berättar att hon säger till studenterna att i muntliga övningar ”är det meningen att föra en ordentlig diskussion, de ska kunna ord och terminologi och tala tillräckligt korrekt så att också en svensk skulle kunna följa med”. Det kommunikativa finns med i hennes tankesätt indirekt, men hon nämner ändå bredden på ordförrådet och korrekthet som exempel på det som hon förväntar sig av en muntlig prestation. Hon säger att hennes bedömning inte var analytisk, utan baserade sig på intuition och erfarenhet. Enligt min tolkning betonar hon organisatorisk kompetens i intervjun.

**Bedömare 6** är ett specialfall i den meningen att hon inte alls har tidigare erfarenhet av att bedöma med CEFR-baserade skalor.

**Tabell 14: Bedömarprofil, bedömare 6**

Erfarenhet som bedömare	Bedömare 6 har inte erfarenhet av bedömning av muntliga prov från tidigare. HY-TALK var första gången hon såg nivåskalan för språkkunskap. Hon har inte fått utbildning i CEFR-skalor.
Reliabilitet	Enligt bedömare 6 fungerar skalan någorlunda bra. Hon säger ändå att hennes bedömning baserade sig på hennes känsla. Hon önskar att hon skulle kunna diskutera med erfarna bedömare om bedömningen för att vara säkrare på sin bedömning.

	<p>Hon tycker att paret i paruppgifter påverkar studentens prestation. Om man är svag jämfört med paret får man lägre betyg, och den som syns mera får ett högre betyg.</p> <p>Hon tycker att det fanns skillnader mellan bedömare mellan t.ex. nivåerna B1 och B2. I helhetsbedömningarna tycker hon att bedömarna var lika, men det fanns variation i vissa uppgifter. Hon hade självt bedömt strängare i någon uppgift och snällare i någon annan jämfört med de andra bedömarna. Hon tycker också att det kan påverka att hon som bedömare hade svenska som modersmål, och alla inte hade det. Hon tycker inte att användningen av skalan var svår, men hon vet inte hur hennes bedömning har lyckats. Hon tycker att diskussion med andra bedömare och utbildning är viktigt.</p>
Validitet	Bedömare 6 tycker att skalan är väl skriven och det var lätt att följa skalan, men att det var svårt att bedöma så många olika delar. Hon skulle hellre ha använt ett enda formulär för hela provet och inte bedömt skilt för alla delmoment.
Autenticitet	Bedömare 6 tycker att i ett provsammanhang kan situationen aldrig bli helt naturlig.
Flyt	Bedömare 6 tycker att vissa uppgifter i HY-Talk var så korta att det var svårt att bedöma flyt på grund av en uppgift. Det var lättare att bedöma en helhetsuppfattning av flyt.
Uttal	Bedömare 6 tycker att man fick en bättre uppfattning av uttalet när eleven hade pratat en längre stund, och önskar att man skulle bedöma uttal bara en gång, inte skilt för varje uppgift i HY-Talk.
Grammatisk korrekthet	Bedömaren problematiserar betydelsen i grammatisk korrekthet jämfört med aktivt muntligt deltagande (eleven talar mycket och inte alltid korrekt) i andra ändan. Vem ska få ett bättre vitsord: den som talar mindre men korrekt, eller den som talar mera men inte lika korrekt? Hon kan inte svara, utan konstaterar att det varierar från fall till fall.
Speciellt för denna bedömare	<p>Bedömare 6 sade att helheten var viktig för henne i bedömningen. Hon tycker att kriterierna och uppgifterna fungerade helt okej, fastän uppgifterna 3 och 4 kunde ha varit en uppgift tillsammans. Hon skulle ha gärna gjort bedömningen som en helhet och inte per uppgift. Å ena sidan säger hon att uppgiftsspecifik bedömning ger möjlighet att lyckas ens i några uppgifter. Å andra sidan kan en helhetsbedömning inte falla på en liten uppgift.</p> <p>Samtalen med en infödd talare var inte helt naturliga, och det var problematiskt att</p>

	<p>hur mycket interlokutören får styra situationen, kommenterade bedömaren.</p> <p>Jämfört med vanliga skolvitsord från 4 till 10 är skalan mera noggrann och det finns exakta kriterier. Bedömaren 6 tycker ändå att det inte har någon betydelse om man har skalan eller den gamla nummerskalan i bedömningen.</p>
--	--

Bedömaren 6 nämner flyt och uttal i sin reflektion om bedömning. Hon säger att hon betonade helheten i bedömningen. Det förblir oklart vad detta betyder. Hon problematiserade naturligheten i provsituationen, vilket jag tolkar som tankar om autenticitet. Hon väger mellan två typer av färdighet: i den ena ändan finns en elev som talar korrekt men inte mycket, och i den andra ändan finns en kommunikativ elev som talar mycket men inte alltid korrekt. Bedömaren väljer inte sida, hon tycker att bedömningen i den situationen varierar från fall till fall. Hon säger: ”jag ger ett plus för en som ändå försöker, men om det låter helt hemskt och det är mycket fel så då drar det ändå till andra hållet”. Hon ger exempel både på bedömningen av drag som hör till organisatorisk kompetens samt på pragmatisk kompetens.

**Bedömaren 7** är en erfaren bedömaren som betonar organisatorisk kompetens i sin bedömning.

**Tabell 15: Bedömarprofil, bedömaren 7**

Erfarenhet som bedömaren	Bedömaren 7 är erfaren som bedömaren, har använt CEFR-baserade skalor mångsidigt och varit med i planerings- och forskningsarbete på national nivå. Av mina informanter har hon haft mest att göra med nivåskalan.
Reliabilitet	<p>Bedömaren 7 betonar att hon skulle ha velat höra eleverna två gånger före bedömningen. Nu blev det en global bedömning som man efteråt försökte differentiera, vilket är åt fel håll enligt bedömaren. Bedömning var svårt eftersom man måste samtidigt fundera på många detaljer.</p> <p>Hon säger att hennes bedömning inte är helt pålitlig, eftersom bedömningen blev dels en kompromiss där hon efteråt tittade på den allmänna trenden och bestämde på grund av den de delar som hon inte hade fyllt i under videon.</p> <p>Hon tror att hon hade bedömt annorlunda om hon hade lyssnat på prestationerna</p>

	<p>en gång till.</p> <p>Hon säger, att pålitligheten i bedömningen kommer från erfarenhet: ”I vanliga arbetet tror jag nog att det är en viss erfarenhet. Jag tror att vi nästan alltid gör en global bedömning som på något sätt sitter i ryggraden. I skalan finns ingredienser som vi tittar samtidigt.” Hon säger, att bedömningen inte kan stämma till hundra procent eftersom alla bedömare är människor och individer. Enligt hennes åsikt brukar lärare med lång erfarenhet bedöma strängare eftersom tidigare kunde studenter bättre svenska. Nuförtiden kan folk inte längre så mycket. Yngre lärare är snällare och kan också läsa skalan som den är. För att få alla på samma linje är utbildning viktigt.</p>
Validitet	<p>Bedömare 7 tycker, att efter att CEFR-nivåskalorna har tagits i bruk har bedömningen blivit mindre felfixerad och mera koncentrerad på balansen. Enstaka fel kan inte fälla en. Hon nämner också, att man kan motivera sin bedömning med nivåskalorna, och använda dem som basis i en analytisk bedömning.</p>
Genomförande av uppgiften	<p>Hon kommenterar texten som handlade om Jere Karalahti: vissa studenter var fullständigt ointresserade av ämnet medan andra hade råkat behandla samma tema på sin kurs, och hade en massa åsikter och färdiga tankar. De förstnämnda lyckades inte visa sina kunskaper i den här uppgiften.</p>
Uttal	<p>”Uttalet och prosodin är vanligen ganska stabilt och ändras inte från uppgift till uppgift. Det är svårt att säga att den biten var bättre i någon uppgift än någon annan. Det var ganska svårt att separera.” Hon tycker att uttalet inte ändras mellan uppgifterna. För henne är uttalet och betoning en speciellt viktig del av muntlig färdighet. Hon tycker att fina skillnader mellan kvalitet på vissa ljud inte är viktigt, men folk måste förstå vad inläraren säger. ”Grunduttalet är väldigt viktigt, och betoningen är viktig för förståelsen.”</p>
Bredd	<p>Studenterna var unistudenter som ska kunna sitt fackspråk, och de ska kunna uttala och betona väl nyckelorden i sitt arbete. Ordförrådets vidd är viktigt för bedömaren, man ska kunna använda undvikande strategier, parafraser, andra ord eller paraplybegrepp. Hon tycker att dessa strategier lär man sig automatiskt när man lär sig sitt modersmål.</p>
Grammatisk korrekthet	<p>Bedömaren betonar också grammatisk korrekthet, men man behöver inte vara perfekt. Hon säger, att man ska kunna reglerna till sextio procent rätt. Man får göra misstag men inte alltid när det är möjligt, inläraren måste i princip veta hur det fungerar. Speciellt på B1-nivån behöver man inte kunna allt väl.</p>
Interaktion/Kommunik	<p>I monologiska uppgifter kan man inte bedöma interaktion. Bedömare talar mycket om interaktion och problematiserar dess utvärdering. Interaktion är svårt att separera från språkbruket, men förmåga att kommunicera och samtala är inte</p>

ation	<p>språkspecifikt. Enligt henne är det fel om man ger pluspoäng av bara kommunikationen: man måste koncentrera sig på språket.</p> <p>Man lär sig interaktionen när man socialiseras i det samhälle man bor i. Det är inte statistiskt men man lär sig det en gång för alla språk. Du är mer eller mindre samma person på alla språk som du talar. I bedömning av språkkunskaper ska man inte bedöma nonverbal kommunikation, utan bara språkspecifik kommunikation. Det är viktigt att det blir kommunikation men det räcker inte om man inte kan språket. Det är inte viktigt att man säger någonting oberoende av vad det är, det borde också vara rätt.</p> <p>Bedömaren 7 tyckte att det var svårt att bedöma språkkunskaper och kommunikation samtidigt, och hon hade betoningen på den lingvistiska delen av språket. Men om man utvärderar språkkunskaperna i ett specifikt språk, så kan inte den här kommunikativa kompetensen som man har i alla språk, som hör till ens sociala kunskap, den kan inte ensam avgöra ens språkkunskaper i svenska. Man ska inte överskatta interaktionen. Det är skillnad mellan språkkunskaper och kommunikativa kunskaper. Man måste i bedömningen ha rätt balans mellan dem. Bedömaren tycker att lärarna i Finland inte har en sådan balans. Man måste också lyckas uttrycka detaljerat, nyanserat innehåll och rikare språk med mera syntax och fraser. Det är bättre än bara prata massor.</p> <p>Bedömaren 7 tycker också att tiden nu är sådan att muntlig kommunikation uppskattas väldigt mycket, och man glömmer lätt att skrift också är kommunikation. Hon tycker också att man borde kräva mera av studenterna, och inte nöja sig med att de talar någonting oberoende av kvalitet.</p> <p>I lärarutbildningen borde man diskutera vad muntlig kommunikation på ett främmande språk är. Det är muntlig kommunikation och språket, inte bara muntlig kommunikation.</p>
Speciellt för denna bedömaren	<p>Bedömaren kritiserar trenden att betona interaktion och kommunikation. Hon tycker att lärarna inte vet vad de borde bedöma när det gäller interaktion på främmande språk. Bedömaren 7 vill betona språkkunskaperna, den lingvistiska delen av språket.</p> <p>Bedömaren 7 definierar också skillnaderna mellan uttal och korrekthet på nivåerna B1 och B2 väldigt noggrant.</p>

Bedömaren 7 skiljer sig mycket från många andra bedömaren i mina intervjuer. Hon kritiserar interaktionens överdrivna roll i det nuvarande bedömningsarbetet (se tabell 14). Hon tycker att interaktion och kommunikation är viktigt, men det är oftast inte språkspecifikt och därför ska man inte låta det påverka för mycket bedömningen av

muntlig språkfärdighet på främmande språk. Hon betonar språkspecifika kunskaper: uttal, prosodi och grammatik. Bedömaren 7 betonar enligt min mening den organisatoriska kompetensen i sin bedömning.

## **5.2 De viktigaste resultaten av intervjuerna**

### **5.2.1 Reliabilitetsfrågan**

Informanterna nämnde i intervjun både problem när det gäller reliabilitet mellan bedömaren och inom en och samma bedömaren. Bedömningsresultaten kan variera enligt mina informanter. Sex bedömaren av sju tycker att det fanns skillnader i bedömningen mellan olika bedömaren som deltog i bedömningssessionerna. Två bedömaren sade att bedömningarna i HY TALK var ganska liknande. Fyra bedömaren nämnde i intervjun att det finns skillnader mellan bedömaren: vissa lärare är snällare och andra strängare och olika bedömaren betonar olika saker.

När det gäller reliabilitet inom bedömaren, fick jag många olika slags kommentarer. Enligt en informant var det problematiskt att bedöma ett utförande (*performance test*), eftersom så många saker sker samtidigt på en kort tid i en kommunikativ situation. Man borde bedöma samtidigt de olika aspekterna i talet, som till exempel grammatik och interaktion. Enligt henne hjälper det om man kan titta på ett utförande på video många gånger. En bedömaren sade att man först bedömer olika eftersom man behöver tid att fokusera på vad det är man skall bedöma. Hon tycker att bedömningen av muntliga språkfärdigheter aldrig blir riktigt rättvis. Några andra informanter förhöll sig lite mera positivt till bedömningens reliabilitet. En bedömaren tycker att hon hade bedömt annorlunda om hon hade lyssnat på prestationerna en gång till. Två bedömaren tog upp elevens synvinkel och därmed validitetsfrågan: de säger att elevens personlighet, "hela människan" är med i bedömningen. Också skalorna togs upp som en problematisk sak i bedömningen. En av informanterna tyckte att skalorna är för grova och därför svåra att använda. Det var också svårt att göra skillnad mellan de olika nivåerna i skalorna. Några informanter nämnde att det fanns oklarheter i de olika nivåerna och faran för att tolka kriterierna på många sätt fanns. Det nämndes också att man måste kunna skalan utantill innan man kan börja använda den i bedömningen. Det kräver tid innan man lär

sig skalans principer och kommer ihåg kriterierna. Fem informanter nämnde utbildningen som ett viktigt sätt att främja skalans användning i skolor. Det är inte lätt att sätta sig in i skalan, utan det kräver utbildning enligt en del av informanterna.

I intervjuerna kom det fram att bedömarna var osäkra på bedömningens enhetlighet. Största delen tyckte att bedömningen inte blev helt enhetlig i bedömningssessionerna. Två tyckte ändå att resultaten blev "ganska liknande". Som en lösning på problemet kom det fram att diskussionen mellan bedömarna efter bedömningen och bedömarens erfarenhet ökar pålitligheten av bedömningen. Bedömarna nämnde också att det lönar sig att titta på elevprestationer på video många gånger för att kunna evaluera olika detaljer.

Kriteriebaserad bedömning togs upp i intervjuerna som en aspekt som kan göra bedömningen enhetligare och på det sättet öka reliabiliteten. Speciellt en av lärarna betonade kriteriebaserad bedömning och dess goda sidor, fast alla informanter nämnde åtminstone att med hjälp av skalan blir bedömningen inte så subjektiv som annars. Nedan finns ett citat av intervjun med lärare 4.

*"Det har nog åtminstone hjälpt mig väldigt mycket. Jag tycker att nu är det mycket enklare att motivera varför man bedömer på ett visst sätt, och kunna visa och ge respons samt motivera varför det blev på ett visst sätt. "*

Tre av informanterna förhöll sig speciellt positiva mot nivåskalan och kriteriebaserad bedömning. De ansåg att skalorna gör bedömningen noggrannare och mera detaljerat. Man kan också lättare motivera sin bedömning och visa studenten vilka saker man tagit till hänsyn. En av lärarna sade också att om alla är på samma nivå enligt skalan, så har läraren gjort ett bra jobb. Skalorna hjälper till att ge rättvisare vitsord enligt några informanter. Informanterna förhöll sig också positivt till en gemensam skala i hela Europa. Den tycktes vara fördelaktig till exempel i studentutbyte.

### **5.2.2 Kommunikation och interaktion i provsituationen**

De största skillnaderna mellan bedömarna gäller förhållningssättet mot elevernas kommunikationskunskaper. En del av informanterna tyckte att kommunikation är det



viktigaste i muntlig språkfärdighet och kan förbättra elevens resultat även om korrektheten inte är så bra, speciellt på lägre nivåerna i skalan. Andra lade inte så stor vikt på kommunikation, utan tyckte att den språkliga delen av språkfärdigheter ska ha den största vikten i bedömningen. En bedömare föreslår att man kunde ha en skild spalt för bedömning av nonverbal kommunikation. En annan bedömare tycker att kroppsspråket inte får påverka bedömningen.

### 5.2.3 Grammatikens roll i muntliga språkkunskaper

För bedömare 1 är det oklart vilka grammatiska strukturer inläraren skall kunna på vissa nivåer i skalan för språkfärdighet. För henne är det inte klart hur man kategoriserar grammatiska strukturer och regler för svenska språket i skalan. För bedömarna 2, 3, 4 och 5 är korrekthet inte så väsentligt om det inte förorsakar brister i kommunikationen. Bedömare 4 konstaterar dock att på C-nivån måste man kunna uttrycka sig rätt och använda rätta grammatiska konstruktioner. Då blir mindre nyanser och idiomatiska uttryck viktigare. Speciellt bedömare 7 tog upp grammatiken som en viktig del av muntliga språkkunskaper. Hon avvek från mängden, eftersom många andra bedömare sade att de inte betonar grammatik. Bedömare 7 definierar också skillnaderna mellan nivåer B1 och B2 väldigt noggrant.

*"Den som får B2 har en rätt stabil prosodi, de vet att ord betonas på första stavelsen. Uttalet av de vanliga fonemen är ganska bra, men inte felfritt. B1 söker ännu väldigt mycket och är osäkrare. Ordförrådet märker man ganska snabbt, om ordförrådet är klart fattigare, så kan det inte bli B2. Signaler, nominalfraser som är helt blandade: Ett vacker dagen och så vidare. Om det är fullständigt inkongruent är det B1. Den som får god kan en gång missa en av de här delarna som i princip är korrekt. B2 ska inte ha alldeles hemska anomalier men man nog får missa i morfologin. Huvudsatsordföljden i princip oftast rätt. Bisatser är inte så väsentliga egentligen."*

De andra bedömarna förklarar inte språkspecifika grammatiska regler lika noggrant. Bedömare 3 nämner, att vissa strukturella fel som stör kommunikationen påverkar hennes bedömning. Hon nämner som exempel pronomen och verbpar som har en fundamental skillnad i betydelse, till exempel *min-din*, *ta-ge* och *köpa-sälja*.

Jag skall i det följande diskutera kortfattat bedömarnas åsikter om grammatik och korrekthet och göra en lös koppling till Eklund Heinonens (2009) arbete om processbarhetens närvaro i muntlig språkfärdighet i svenska som främmande språk. I Tisusprovet befann sig största delen av de underkända på nivå 3 i processbarhetshierarkin, och de flesta godkända kunde behärska strukturer på frasnivå (nivå 4 i PT-hierarkin). De kommentarer som jag fick från bedömarna är inte i samma linje med Tisusresultaten. Bedömare 3 nämner enstaka ord som kan vara avgörande i bedömningen av korrektheten. Detta skulle enligt processbarhetsteorin finnas på nivå 1 i hierarkin. Bedömare 7 beskriver nivå B1, som alltså är godkänd på universitetsnivå i Finland (se tabell 2 på sida 14), med att inkongruenta nominalfraser kan förekomma. Om inläraren använder uttryck *ett vacker dagen\**, sätter bedömare 7 denna på nivå B1. Först på nivå B2 nämner hon att huvudsatsordföljden borde vara oftast rätt. Bisatser tycker hon att är inte väsentliga i muntligt språkbruk. Dessa kopplingar är förstås inte väsentliga, men ger en uppmuntring för vidare studier i kopplingar mellan processbarhetshierarkin och nivåskalan för språkkunskaper. När det gäller mina informanter, beskrev de åtminstone inte språkkunskaperna med hjälp av de drag som är viktiga i processbarhetsteorin när det gäller svenska språket. Om man skulle se positiva kopplingar mellan CEFR-skalornas nivåer och processbarhetshierarkin, skulle de kanske kunna användas som beskrivande exempel för svenskan i en språkspecifik nivåskala i framtiden. En sak är nämligen klar: det finns ett behov för språkspecifika beskrivningar av grammatiska konstruktioner i svenska i nivåskalan för språkfärdighet.

### **5.3 Reliabilitet av bedömning enligt nivåskalan för språkfärdighet**

I detta kapitel skall jag svara på forskningsfråga nummer två. I följande analys skall jag alltså behandla reliabiliteten mellan bedömarnas bedömningsresultat i projektet HY-TALK för svenska språkets del. Resultaten presenteras i två delar: först går jag igenom gymnasieelevernas, och sedan högskolestudenternas bedömningsresultat. Av tabell 16 nedan ser man lärarnas helhetsbedömning av gymnasieelevernas utföranden.

**Tabell 16: Helhetsbedömning, gymnasienivå**

Elev nr	Bedömare 1	Bedömare 2	Bedömare 3	Bedömare 4
1	A2.2+	A2.2	A2.2	A2.2
2	A2.1	A2.1	A2.1	A2.2
3	A2.2	A2.1	A2.2	A2.2
4	A2.1	A2.2	A2.2+	B1.1
5	A2.1	A2.1	A2.1	A2.1
6	A2.1-	A2.1	A2.1	A2.1
7	A2.2++	A2.2+,B1.1	B1.2	B1.1
8	A2.1	A2.1	B1.2	A2.1
9	A2.1-	A2.1	A 1.2	A1.3
10	A1.3	A1.3	A 1.2	A1.3
11	A2.1-	A2.1		A2.1
12	A1.3	A1.3		A2.1

Bedömningen på gymnasienivå var för det mesta enhetlig, vilket kan konstateras på basis av tabellen. Det fanns dock vissa skillnader mellan bedömarna. Lärarna 3 och 4 gav lite högre poäng än de andra, och använde skalan bredare. När det gäller elev nummer 4 är en viss spridning synlig. Lärare 1 har satt henne på nivå A2.1, medan lärare 2 och 3 på nivå A2.2, och lärare 4 på nivå B1.1. Eleverna 7 och 8 har också fått betydligt högre vitsord av lärare 3 jämfört med de andra bedömarna. 7 och 8 var ett par och gjorde uppgifterna alltså tillsammans.

För att ytterligare kontrollera mina resultat har jag utfört statistiska t-test för att jämföra om resultaten skiljer sig från varandra statistiskt sett. För att kunna jämföra de olika bedömarnas bedömningsresultat statistiskt har jag omvandlat bedömningsresultaten till siffror enligt följande mönster (se tabell 17 nedan). Plus- eller minustecken (som till exempel i A2.2+) har inte beaktats.

**Tabell 17: Nivåskalorna i siffror**

Nivåskalorna omvandlade till siffror för det statistiska t-testet			
A1.1	1	B1.1	6
A1.2	2	B1.2	7
A1.3	3	B2.1	8
A2.1	4	B2.2	9
A2.2	5	C1.1	10

Resultaten blev följande. Med 95% statistisk signifikans kan man säga att bedömningen var enhetlig vad gäller eleverna nummer 1, 5 och 6. Med andra ord är 33% av

elevbedömningarna på gymnasienivå helt entydiga (alla 4 bedömningarna är lika). Statistiskt sett finns det ändå variation i största delen av elevbedömningarna. Tolkningen av resultaten är trots detta inte entydig, eftersom samplen endast täcker 4 olika bedömare. Därför måste man förhålla sig försiktigt till resultaten. För att nå statistiskt trovärdiga resultat borde samplet täcka minst 30 bedömare (Gurajati 2003). I det nuvarande fallet tillåter testet inte alls variation.

**Tabell 18: Statistiskt test av resultaten på gymnasienivå**

Gymnasienivå Elev	Bedömningsresultatet med 95% statistisk signifikans mellan värden			Bedömningens enhetlighet (JA/NEJ)
	Minst	Högst	Omvandlat till nivåskalan	
1	5,000	5,000	A2.2	<b>JA</b>
2	3,556	4,944	A2.1 - A2.2	NEJ
3	4,056	5,444	A2.1 - A2.2	NEJ
4	3,867	6,133	A2.1 - B1.1	NEJ
5	4,000	4,000	A2.1	<b>JA</b>
6	4,000	4,000	A2.1	<b>JA</b>
7	4,421	7,079	A2.1 - B1.2	NEJ
8	2,668	6,832	A1.3 - B2.1	NEJ
9	1,921	4,579	A1.2 - A2.2	NEJ
10	2,056	3,444	A1.2 - A1.3	NEJ
11	4,000	4,000	A2.1	<b>JA</b>
12	2,532	4,135	A1.3 - A2.1	NEJ

I det följande behandlas bedömningsresultaten av helhetsbedömningen på högskolenivå. På basis av tabell 19 nedan kan det konstateras att bedömningen av högskolestudenter är ganska enhetlig med vissa undantag.

**Tabell 19: Helhetsbedömning, högskolenivå**

Elev	Bedömare 5	Bedömare 6	Bedömare 7	Bedömare 1	Bedömare 4
1	B2.2	B2.1	B2.2	B2.1	B2.1
2	B2.1	B1.1	B1.2	B1.2	B1.2
3	B2.1	B1.2	B2.1	B2.1	B1.2
4	B2.1	B1.2	B1.2	B2.1	B1.2
5	B1.2	B1.2	B1.2	B1.2	B1.1
6	B1.1	B1.1	B1.1	B1.2	B1.1
7	B1	A2.2	B1.1	A2.2	B1.1
8	A2	A2.2	A2.2	A2.1	A2.2

9	B2	A2.2	B2.1	B1.1	B1.1
10	B1	A2.1	B1.2	A2.2	A2.2

Bedömning av studenterna 7 och 8 samt av 9 och 10 har en liten spridning. Jag vill relatera resultaten till tabell 4 i avsnitt 3.5, där det kommer fram att studenten borde finnas på B-nivå för att kunna få godkänd i så kallad tjänstemannasvenska vid en högskola i Finland. Jag tycker att skillnaderna i bedömningen är betydelsefulla om vitsorden av en och samma student varierar mellan A och B-nivåerna. Detta betyder att studenten inte skulle ha blivit godkänd av vissa bedömare medan hon skulle ha fått godkänd av andra bedömare. Studenterna 7, 9 och 10 har fått vitsord som skiljer sig på det ovannämnda sättet. Student nummer 7 har fått vitsordet B1.1 (B1 från bedömare 5) från 3 bedömare, samt A2.2 från bedömarna 6 och 1. Student 9 har fått A2.2 från bedömare 6. Student 10 fick vitsord på B-nivå från bedömare 5 och 7, men de andra bedömarna har gett studenten vitsord A2.2. Bedömare 6 har till och med gett student 10 vitsordet A2.1.

Jag vill också ta upp skillnaderna i bedömningsresultaten mellan bedömarna 5 och 7 (med mer kognitiv syn på muntliga språkkunskaper och betoningen på den organisatoriska kompetensen) och 1 och 4 (med mera kommunikativ syn på muntlig språkfärdighet, fokus på den pragmatiska kompetensen). Skillnaderna i bedömningen mellan dessa två bedömartyper är inte stora men bedömarna 5 och 7 har ofta gett lite högre vitsord till eleverna. Skillnaden kan ses tydligt speciellt när det gäller eleverna nummer 9 och 10. Bedömare 6, som hade minst erfarenhet av bedömning av muntliga språkfärdigheter i svenska enligt nivåskalan för språkfärdighet, har varit ganska strikt i bedömningen. Hon har oftast gett studenterna lite lägre vitsord jämfört med de andra bedömarna. Hon skulle inte ha gett godkänd (på högskolenivå i finska högskolor) till studenterna 7, 9 och 10, medan dessa studenter skulle ha fått godkänt vitsord från de mera kognitiva bedömarna.

Det statistiska kontrolltestet för högskolestuderandenas bedömningar ger inte uppmuntrande resultat. Ingen av resultaten är statistiskt sett entydig, utan det finns variation i bedömning av varje studerande. Här måste det igen konstateras att detta

statistiska resultat är av lågt värde eftersom samplet är endast 5 bedömare per studerande.

**Tabell 20: Statistiskt test av resultaten på högskolenivå**

Högskolenivå Elev	Bedömningsresultatet med 95% statistisk signifikans mellan värden			Bedömningens enhetlighet (JA/NEJ)
	Minst	Högst	Avrundat till nivåskalan	
1	7,640	9,160	B2.1 - B2.2	NEJ
2	6,019	7,981	B1.1 - B2.1	NEJ
3	6,840	8,360	B1.2 - B2.1	NEJ
4	6,640	8,160	B1.2 - B2.1	NEJ
5	6,179	7,421	B1.1 - B1.2	NEJ
6	5,579	6,821	B1.1 - B1.2	NEJ
7	4,840	6,360	A2.2 - B1.1	NEJ
8	3,840	5,360	A2.1 - A2.2	NEJ
9	4,738	8,462	A2.2 - B2.1	NEJ
10	3,817	6,983	A2.1 - B1.2	NEJ

Svaret på min forskningsfråga nummer 2 gällande reliabilitet av bedömningsresultaten blir att eftersom det finns skillnader i bedömningsresultaten, blir reliabiliteten inte perfekt i min fallstudie, men den är ändå relativt hög kvalitativt tolkat.

## 6 Diskussion av resultaten

Det fanns många problem som bedömarna nämnde i intervjuerna. En del av problemen gällde bedömning av muntliga språkfärdigheter allmänt, och en del gällde användningen av skalan i läroplansgrunderna. Största delen av informanterna nämnde att det finns skillnader i bedömningen mellan olika bedömare och också inom en och samma bedömare i olika bedömningssituationer. En av informanterna ifrågasatte också skalornas validitet. Hon tyckte att skalorna inte används rätt och att de inte mäter rätta saker.

Skillnaderna i förhållningssättet till interaktion och kommunikation utgör i min mening de mest intressanta resultaten av intervjuerna. Bedömarnas attityd till bedömning av de kommunikativa aspekterna av muntlig språkfärdighet skiljde sig drastiskt från varandra. Detta resultat är i och för sig ingenting nytt, men tyder på de problem som finns i bedömningen av muntliga färdigheter i svenska i Finland.

Några av bedömarna talade om den organisatoriska kompetensen (till exempel uttal och grammatik) i intervjusituationen, andra betonade den pragmatiska kompetensen (till exempel undvikande strategier och nonverbal kommunikation). Det kom tydligt fram att bedömarna lade märke till olika delar av den kommunikativa språkförmågan i sin bedömning. I denna syn fanns det stora skillnader mellan lärarna. Jag tycker att detta kan förorsaka problem i bedömningen av muntliga språkfärdigheter. Synen på vad språkkunskaperna i svenska innefattar måste påverka bedömarens konkreta bedömningsarbete. En del av mina informanter sade att kommunikation och till och med nonverbal kommunikation är en viktig del av språkkunskaperna och skall tas med i bedömningen. Två bedömare tyckte att nonverbal kommunikation är någonting som inte skall påverka bedömningen av muntliga språkfärdigheter. Skillnaderna i de två grundläggande synerna på vad muntlig språkfärdighet innehåller är enligt min tolkning stora mellan dessa två grupper av informanter. Man måste ändå påpeka, att dessa skillnader också kan bero på bedömningens karaktär. Bedömningen av gymnasieelever var en allmän bedömning av språkkunskaper i svenska språket. Bedömningen av högskolestudenter var å sin sida en yrkesspecifik bedömning där kunskaper väsentliga för det framtida yrket kan ha betonats. Det fanns skillnader i utgångspunkterna i bedömningen under de två olika bedömningssessioner, och inte endast mellan olika bedömare.

Det kunde inte ses några stora skillnader i bedömningsresultaten mellan de två typerna av bedömare. De som inte betonar kommunikation i sin bedömning hade ändå gett några studenter lite högre vitsord jämfört med de mera kommunikationsbetonande bedömarna. Denna skillnad kan också förklaras med informanternas erfarenhet: de förstnämnda är vana vid att bedöma elever på just den nivån som högskoleeleverna i HY-TALK var. De som betonade kommunikation i större utsträckning å andra sidan är inte så vana vid att bedöma högskolestudenter på dessa nivåer i skalan. Erfarenhet i yrkesbetonad bedömning kan i det här fallet ha mera betydelse än den allmänna synen på

språkkunskaper. För att kunna dra trovärdiga slutsatser på basis av de statistiska testen, borde man ha ett större sampel av elevbedömningarna.

I min undersökning tycks det inte finnas stora skillnader i bedömningen mellan de två grupperna, fastän deras syn på bedömningen av muntliga språkkunskaper i svenska skiljer sig betydligt från varandra. När det gäller mitt material, kan det konstateras att bedömningens reliabilitet var relativt bra, och inga betydligt stora skillnader i bedömningen fanns. Trots detta fanns det några skillnader i bedömningen om man relaterar resultaten till kriterierna för att få godkänd i "tjänstemannasvenska" på högskolenivå i Finland. Tre elever av tio skulle ha fått godkänd av en del bedömare medan andra inte skulle ha godkänt dem. Det måste också påpekas, att mitt material är så litet att inga generaliserande slutsatser kan dras.

Mina resultat tyder på att fastän reliabiliteten av bedömningsresultaten är relativt bra i HY-TALK-bedömningen (Hildén & Martikainen 2009), och fastän nivåskalan har validerats med kvantitativa tester, finns det skillnader i bedörnarnas betoningar i bedömningen. Bedörnarna använde samma kriterier, men betonade olika delar av den kommunikativa språkförmågan i sin bedömning. Detta kan anses minska testets validitet kvalitativt sett. Fastän reliabiliteten var bra, kom det fram problem gällande testets validitet på grund av olika tolkningar av bedömningskriterierna.

## **6.1 Hur borde man gå vidare?**

I litteraturen kring språkbedömning betonas ofta utbildning som ett sätt att förbättra bedömningens kvalitet: bättre reliabilitet och validitet. Till exempel Alderson et al (1995: 126) konstaterar, att bedömarutbildning är ytterst viktigt, och institutionerna som är ansvariga för den borde också reservera tillräckligt med tid för utbildningen i att använda en skala. Bedörnarna måste bli bekanta med evalueringssystemet, schema eller skala, och lära sig tillämpa det konsekvent. De måste också få veta vad de borde göra i en oväntad situation. (ibid. 105) Också erfarna bedömare behöver ständig uppdatering och tilläggsutbildning. Bedömarutbildning är också avgörande när det gäller provets validitet och reliabilitet. Utan välutbildade, konsekventa bedömare kan ett prov inte vara



pålitligt. En utmaning i subjektiv bedömning är att bedömarna förstår principerna bakom skalan de använder och kan tolka beskrivningarna på de olika nivåerna i skalan konsekvent. När det gäller muntlig bedömning, kan bedömarna också behöva utbildning i interaktionen med deltagarna i bedömningssituationen. Standardisering är också viktigt för att bedömarna ska få en bild av olika typer av utföranden.

Alderson et al (1995: 112) menar, att också erfarna bedömare måste delta i utbildning eftersom reliabiliteten kan minska när en erfaren bedömare blir självbelåten (*complacent*) i sitt arbete. Utbildningen kan ske till exempel så, att bedömarna lyssnar på utföranden från en videoinspelning. De borde ges möjlighet att lyssna många gånger på ett utförande och sedan ge en individuell bedömning. Efter det behövs diskussion av poängsättningen, finjustering och till sist uppnående av ett samförstånd i poängsättningen. Institutionen som ordnar testerna borde följa en linje för hur stor grad av enighet de förutsätter av bedömarna och bedömarna borde uppnå en viss standard innan de ges lov att bedöma på riktigt.

Också enligt min åsikt är utbildning och fortbildning ytterst viktigt, men också skalorna behöver anpassning. De är inte lätta att börja använda, och därför har bedömarna svårt att motivera sig själva att studera CEFR-skalorna och skalan i läroplansgrunderna. I Finland har man modifierat den europeiska nivåskalan med en ännu finare indelning där det finns tio olika nivåer i stället för sex i CEFR. Å ena sidan gör detta bedömningen lättare, men å andra sidan också ännu mera komplicerad, speciellt för en nybörjare.

Jag tycker att utbildning och skalornas anpassning till egna behov borde gå hand i hand. Att kunna modifiera och anpassa skalorna enligt sina egna behov gör dem också mera användbara för lärarna. Modifieringen borde också ske så att det inte medför sämre reliabilitet. I Finland har man redan varit modig i att modifiera skalan för de finska läroplansgrunderna. Jag tycker att man borde vara ännu modigare och lista mera konkreta exempel på vad kriterierna omfattar i de olika språken och på olika nivåer inom språken. I läroplansgrunderna finns det exempel på grammatikkunskaper som visar läraren vad som anses höra till de olika nivåerna i skalan. Språkexperterna i de olika

språken inom förvaltningen, universiteten och utbildningsstyrelsen och undervisningsministeriet behövs för att ge konkreta råd till lärarna i form av till exempel en handbok till nivåskalans användning. På Utbildningsstyrelsens nätsida för lärare, edu.fi, finns det en bra artikel om nivåskalans användning (Hildén 2008), men också en mera praktiskt orienterad och enkel handbok skulle vara till nytta för många lärare.

Det är viktigt att tydligt definiera för bedömarna vilka delar av den kommunikativa språkförmågan som skall tas med i bedömningen. På testnivå kan man besluta om det är den organisatoriska kompetensen, den pragmatiska kompetensen eller båda kompetenserna som bedömarna skall beakta i bedömningen. Detta görs säkert i praktiken när språktesten planeras, men åtminstone i min fallstudie kom det fram att det fanns stora skillnader i betoningarna i bedömningen mellan olika bedömare. De använde samma skala som kriterier för bedömning, men det fanns några skillnader i bedömningen. För att öka reliabiliteten samt validiteten för ett språktest, är det viktigt att tydligt informera bedömarna om de kompetenser som man skall lägga märke till just i det språktestet.

Det är viktigt att komma ihåg att CEFR inte enbart är ett pedagogiskt verktyg och inte heller bara ett språkvetenskapligt hjälpmedel, utan ett språkpedagogiskt redskap. Det är viktigt att språkvetarna samarbetar i modifieringen och tillämpningen av skalan. Bara med hjälp av språkspecifikt men samtidigt reliabelt definierade nivåer i skalan kan man börja använda nivåerna i det verkliga skollivet. Det är viktigt att göra skalorna så tydliga som möjligt för att lärarna skulle bli motiverade till att börja använda dem. Ju snabbare man förstår skalornas logik och litar på sig själv som deras användare, desto positivare blir lärarnas syn på skalor. Om bedömarna, alltså oftast lärarna, förhåller sig positivt till skalorna, används de kanske också med större sannolikhet.

## 7 Avslutning

CEFR-skalan och den finländska nivåskalan för språkfärdigheter har snabbt blivit en viktig del av språkbedömningen i den grundläggande utbildningen samt gymnasieutbildningen i Finland. Kriteriebaserad bedömning har varit välkommen i Finland och den gör för sin del bedömningen mera reliabel och objektivare.

Bedömningen av muntliga språkfärdigheter ställer sina egna speciella utmaningar till bedömningen. När man bedömer en muntlig språkprestation, måste man ta hänsyn till många olika saker samtidigt. En del av mina informanter ifrågasatte bedömningens karaktär: är det rätt att bedöma när en elev uppträder? Kan man bedöma språkfärdigheter utan att bedöma en persons personlighet? När man bedömer muntlig färdighet tar man lätt hänsyn till personens kommunikationsförmåga, gester och mimik, livlighet och motivation. Grammatik blir ofta mindre viktig och interaktionen betonas i bedömningen. En del av informanterna tyckte att kommunikation är en viktig del av muntlig språkfärdighet och måste därför tas med i bedömningen. Några tyckte ändå att den nonverbala delen av kommunikationskunskaper inte får påverka bedömningen.

Själva nivåskalan kritiserades också. Kriterierna i skalorna tycktes vara otydliga. Jag tycker att en gemensam kriterieskala för alla språk lätt kan bli otydlig och för allmän för att kunna användas i praktiken. Därför är det viktigt att skalorna tillämpas till de olika språkens behov och med hjälp av konkreta språkliga exempel görs det lättare för lärarna att tillägna sig skalan. Lärarna behöver modellbedömningar och konkreta exempel på grammatiska och kommunikativa språkspecifika aspekter som visar hur skalan fungerar i praktiken. Utbildning spelar en avgörande roll för hur snabbt och djupt skalan tillägnas och börjar användas i skollivet.

Mitt material är för litet för att man ska kunna dra några allmänna slutsatser av det, men det ger ändå en viss bild av de problem som finns i bedömningen av muntlig färdighet enligt nivåskalan. Muntlig språkfärdighet är svårt att definiera och begränsa. Enligt min åsikt är den kommunikativa sidan och interaktionen en viktig och väsentlig del av

mundlig språkfärdighet som kan och borde beaktas också i bedömningen av muntliga språkfärdigheter. CEFR och den finländska nivåskalan betonar inte interaktion.

Muntliga språkfärdigheter är krävande och utmanande att bedöma, och därför är det också svårt att formulera bra fungerande allmänna skalor som skulle täcka alla språkliga krav och definiera muntliga språkfärdigheter genomgripande. En lösning kunde vara att man bestämmer vilka aspekter av muntlig språkfärdighet man tar med i bedömningen och vad man lämnar utanför. För att kunna upprätthålla bra reliabilitet, måste man ha gemensamma spelregler. Man behöver inte bestämma att detta gäller hela Europa, men till exempel i de nationella språkproven måste man ha gemensamma regler för hur man bedömer till exempel interaktionen i ett muntligt utförande. Här är nivåskalan i viss mån otydlig: den tar inte ställning, den bara låter bli att svara på frågan.

Det är inte lätt att ta ställning till hur interaktionen borde bedömas i muntliga språkprov. Olika språkteoretiker har olika syn på de grundläggande principerna i språkbedömningen. Som det kom fram i intervjuerna, finns också skillnader i praktiken. CEFR-skalan har inte kunnat gå genvägar och bestämma ett enda förhållningssätt på interaktionsfrågan. Så länge som frågan om hur muntlig språkfärdighet borde bedömas förblir utan svar, varierar också bedömningen.

## Litteratur

- Alderson, J. Charles, Clapham, Caroline & Wall, Dianne (1995): *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Bachman, Lyle F. & Palmer, Adrian S. (1996): *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Bachman, Lyle F. (1990) *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bausch & Christ & König (2003): "Der Gemeinsame europäische Referenzrahmen für Sprachen: Zustimmung, aber...!" I: Morrow. Keith (ed.): *Insights from the Common European Framework*. Oxford: Oxford University Press.
- Brown, Annie (2004): Discourse analysis and the oral interview: Competence or performance? I: Boxer, Diana & Cohen, Andrew D (eds.), *Studying speaking to inform second language learning*. England: Clevedon. S. 243-282.
- Brown, Annie & Lumley, Tom (1997): Interviewer Variability in Specific-Purpose Language Performance Tests. I: Huhta, Ari et al. (eds.) (1997): *Current Developments and Alternatives in Language Assessment - Proceedings of LTRC 96*. Jyväskylä: University of Jyväskylä.
- Brown, H. Douglas (2004): *Language Assessment: Principles and Classroom Practices*. Pearson Education. New York.
- Carlsen, Cecilie (2003): Guarding the guardians. Rating scale and rater training effects on reliability and validity of scores of and oral test of Norwegian as a second language. Nordisk institut. Universitetet i Bergen.
- CEFR: Common European Framework of Reference for Languages: Learning, Teaching, Assessment (2001). Council of Europe. Strasbourg Cedex. Tillgängligt på: [http://www.coe.int/T/DG4/Linguistic/Source/Framework\\_EN.pdf](http://www.coe.int/T/DG4/Linguistic/Source/Framework_EN.pdf) (Hämtat 2.2.2009)
- Davies, Alan (Ed.) et al. (1999): *Studies in Language Testing 7. Dictionary of language testing*. Cambridge: Cambridge University Press.
- Eklund Heinonen, Maria (2009): Processbarhet på prov. Bedömning av muntlig språkfärdighet hos vuxna andraspråksinlärare. Institutionen för nordiska språk vid Uppsala: Uppsala universitet.
- Elsinen, Raija & Juurakko-Paavola, Taina (2006): *Korkeakouluopiskelijoiden ruotsin kielen taidon arviointi*. Tavastehus: Hämeenlinnan ammattikorkeakoulu.
- Eurooppalainen viitekehys. Kielten oppimisen, opettamisen ja arvioinnin yhteinen eurooppalainen viitekehys (2003). Helsingfors: WSOY.
- Europeisk referensram för språk (2007): *Gemensam europeisk referensram för språk: lärande, undervisning och bedömning*. Stockholm: Skolverket.

- Green-Vänttinen, Maria & Rosenberg-Wolff, Carita (2007): En svensklärare måste kunna tala god svenska. Pedagogiskt utvecklingsarbete för undervisning i och utvärdering av muntlig färdighet. I: Lehti-Eklund, Hanna (red.) 2007: Att växa till lärare – svensklärarutbildning i utveckling. Nordica. Institutionen för nordiska språk och nordisk litteratur. Helsingfors: Helsingfors Universitet.
- Grunderna för läroplanen för den grundläggande utbildningen (2004). Helsingfors: Utbildningsstyrelsen.
- Grunderna för läroplanen i gymnasieutbildning för ungdomar (2003). Helsingfors: Utbildningsstyrelsen.
- Gurajati (2003): Basic econometrics. Fjärde upplaga. New York: MacGraw-Hill Higher Education.
- Hildén, Raili & Martikainen, Marja K. (2009a): Arvioijapalautetta puhekokeesta – saksan kielen arvioijaraadin näkemykset validiusargumenttia rakentamassa. Helsinki: Helsingin yliopisto.
- Hildén, Raili & Martikainen, Marja K. (2009b): Rater feedback on speaking assessment. EALTA Conference 4.-7. June 2009. Opublicerad. Helsingfors universitet.
- Hildén, Raili & Takala, Sauli (2007): Relating Descriptors of the Finnish School Scale to the CEF Overall Scales for Communicative Activities. I: Koskensalo, Annikki, Smeds, John et al. (ed.) Foreign languages and multicultural perspectives in the European context = Fremdsprachen und multikulturelle Perspektiven im europäischen Kontext. Berlin: Lit Verlag.
- Hildén, Raili (2008): Taitotasot opetuksen suunnittelun apuna. Utbildningsstyrelsen. Tillgängligt på: <http://www.edu.fi/page.asp?path=498,1329,1513,29698,29944> (Hämtat 15.6.2008)
- Huttunen, Irma & Takala, Sauli (2004): Euroopan Neuvoston toiminta kielikoulutuksen kehittämiseksi ja sen vaikutus kielisuunnitteluun Suomessa. I: Sajavaara, Kari & Takala, Sauli (red.): Kielikoulutus tienhaarassa. Jyväskylä: Jyväskylän yliopisto.
- Hämäläinen, Leena, Väisänen, Tuula & Lomama, Sirkku (2007): Perus- ja toisen asteen kielikoulutus: peruskielitaidosta monikielisyyteen? I: Pöyhönen, Sari & Luukka, Minna.-Riitta: Kohti tulevaisuuden kielikoulutusta. Kielikoulutuspoliittisen projektin loppuraportti. Jyväskylä: Jyväskylän yliopisto. Soveltavan kielentutkimuksen keskus.
- Kaftandjieva, Felianka & Takala, Sauli (2002): Council of Europe Scales of Language Proficiency: A Validation Study. I: Anderson, Charles (red.): Common European Framework for Reference for Languages: Learning, teaching, assessment. Case Studies. Strasbourg: Council of Europe. Tillgängligt på: [http://www.coe.int/T/DG4/Portfolio/documents/case\\_studies\\_CEF.doc](http://www.coe.int/T/DG4/Portfolio/documents/case_studies_CEF.doc) (Hämtat 2.1.2008)
- Kalliomäki, Heidi (2007): "I find that I am both enthusiast and critic": how teachers experience spoken language assessment through the Common European Framework. Opublicerad pro gradu-avhandling. Helsingfors: Helsingfors universitet.
- Kvale, Steinar (1997): Den kvalitativa forskningsintervjun. Lund: Studentlitteratur.
- Lewkowicz, Jo A. (1997): Authentic for whom? Does Authenticity Really Matter? I: Huhta, Ari et al. (eds.): Current Developments and Alternatives in Language Assessment - Proceedings of LTRC 96. Jyväskylä: Jyväskylä universitet.
- Lindberg, Inger (1997): Lärarhandledning. Nationellt prov för sfi: 3. Stockholm: Skolverket. Liber.
- McNamara Tim (2000): Language Testing. New York: Oxford University Press.

- McNamara, Tim & Roever, Carsten (2006): *Language Testing: The Social Dimension*. Language Learning Research Club, University of Michigan: Blackwell Publishing.
- McNamara, Tim (1996): *Measuring second language performance*. London och New York: Addison Wesley Longman Limited.
- Milanovic, M (2002): *Common European Framework for Reference for Languages: Learning, teaching, assessment. Language examining and test development*. Council of Europe. Strasbourg Cedex. Tillgängligt på:  
<http://www.coe.int/T/DG4/Portfolio/documents/Guide%20October%202002%20revised%20version1.doc> (Hämtat 2.1.2008)
- Mäkelä, Klaus (red.) 1990: *Kvalitatiivisen aineiston analyysi ja tulkinta*. Helsingfors: Gaudeamus.
- North, Brian (1996): *The Development of a Common Framework Scale of Descriptors of Language Proficiency Based on a Theory of Measurement*. S. 423-445. I: Huhta, Ari & Kohonen, Viljo et al.: *Current Developments and Alternatives in Language Assessment – Proceedings of LTRC 96*. Jyväskylä: Jyväskylän universitet och Tampere universitet.
- Pöyhönen, Sari & Luukka, Minna.-Riitta (2007): *Kohti tulevaisuuden kielikoulutusta. Kielikoulutuspoliittisen projektin loppuraportti*. Jyväskylä: Jyväskylän yliopisto. Soveltavan kielentutkimuksen keskus.
- Reed, Daniel J, & Cohen, Andrew D. (2001): *Revisiting raters and ratings in oral language assessment*. I: Elder, C. (Ed.) et al.: *Studies in Language Testing 11*. Cambridge: Cambridge University Press.
- Starr Keddle, Julia (2004): *The CEF and the secondary school syllabus*. I: Morrow, Keith (ed.): *Insights from the Common European Framework*. Oxford: Oxford University Press.
- Strömquist, Siv (1998): *Uppsatshandboken*. Uppsala: Hallgren & Fallgren Studieförlag AB.
- Tarnanen, Mirja, Huhta, Ari & Pohjala, Kalevi (2007): *Mitä on osaaminen? Kielitaidon arviointi vastaajana*. I: Pöyhönen, Sari & Luukka, Minna.-Riitta (2007): *Kohti tulevaisuuden kielikoulutusta. Kielikoulutuspoliittisen projektin loppuraportti*. Jyväskylä. Jyväskylän yliopisto. Soveltavan kielentutkimuksen keskus.
- van Ek, J. A. och Trim, J. L. M. (1990) *Threshold Level 1990*. Strasbourg: Europarådet.
- Wrigstad Thomas (2004): *Sfi-provet som bedömningsinstrument – teoretiska och praktiska perspektiv*. I: Hyltensam, Kenneth och Inger Lindberg (red.), *Svenska som andraspråk – i forskning, undervisning och samhälle*. Lund. Studentlitteratur. S. 715-742.

## Bilaga 1: Intervjufrågor

Intervjuaren presenterar kort vad intervjun handlar om före man börjar. Hon har med sig CEF och skolskalan som kopia samt evalueringsblanketten som informanten har fyllt i tidigare i Hy-talk-evalueringen.

Informantens ålder, utbildning, kön, nuvarande position

1. Har du använt skolskalan eller CEF i ditt arbete tidigare?
2. Hur bekant är du med CEF och skolskalan i läroplansgrunder?
3. Vad tycker du om användningen av CEF och skolskalan i läroplansgrunder i Finland och i Europa (allmänt)?
4. Hur lyckades du i din åsikt i evalueringen på torsdagen 10.1.2008 (HY-talk-pilotevaluering)?
5. Hur upplever du bedömningen av (speciellt) muntliga språkfärdigheter enligt skolskalan?
6. Vad anser du om kriterierna i skalan, speciellt gällande muntlig färdighet? Vad är positivt? Vad är negativt?
7. Finns det problem i bedömningen av muntliga språkfärdigheter enligt skalan och dess kriterier?
8. (Om svar på 7 är positivt) Hur kunde eventuella problem i bedömningen enligt skalan lösas i din åsikt?
9. (Om svar på 7 är positivt) Kunde man förbättra eller tillämpa skalan så att problem skulle lösas



## Bilaga 2: Skalan för muntlig färdighet i läroplansgrunderna

Täitotaso A1		Suppea viestintä kaikkein yksinkertaisimmissa tilanteissa		
	Kuullun ymmärtäminen	Puhuminen	Luettun ymmärtäminen	Kirjoittaminen
A1.3 Toimiva alkiesitel itaito	<ul style="list-style-type: none"><li>* Ymmärtää yksinkertaisia lausumia (henkilökohtaisia kysymyksiä ja jokapäiväisiä ohjeita, pyyntöjä ja kieltoja) rutiiniaisissa keskusteluissa tilanneyhteyden tukenana.</li><li>* Pystyy seuraamaan yksinkertaisia, välittömiin tilanteisiin tai omaan kokemukseen liittyviä keskusteluja.</li><li>* Yksinkertaisenkin viestin ymmärtäminen edellyttää normaalia hitaampaa ja kuuli-jalle kohdennettua yleiskielistä puhetta.</li></ul>	<ul style="list-style-type: none"><li>* Osa kertoo lyhyesti itsestään ja lähipiiristään. Selvittää kaikkien yksinkertaisimmista vuoropuheluista ja palvelutilanteista. Tarvii joskus puhelukumppanin apua.</li><li>* Kaikkein tutuimmat jaksot sujuvat, muualla tauot ja katkokset ovat hyvin ilmeisiä.</li><li>* Ääntäminen voi joskus tuottaa ymmärtämisongelmia.</li><li>* Osa rajallisen joukon lyhyitä, ulkoa opeteltuja ilmauksia, keskeisintä sanastoa ja perustason lauserakenteita.</li><li>* Alkeellisessäkin puheessa esiintyy paljon peruskiehtiopivirheitä.</li></ul>	<ul style="list-style-type: none"><li>* Pystyy lukemaan tuttuja ja joitakin tuntemattomia sanoja. Ymmärtää hyvin lyhyitä viestejä, joissa käsitellään arkielämää ja rutiinipahutunutta tai annetaan yksinkertaisia ohjeita.</li><li>* Pystyy löytämään tarvitsemansa yksittäisen tiedon lyhyestä tekstistä (postikortti, säätiedotukset).</li><li>* Lyhyenkin tekstipätkän lukeminen ja ymmärtäminen on hyvin hidasta.</li></ul>	<ul style="list-style-type: none"><li>* Selvittää kirjoittamalla kaikkein tutuimmista, helposti ennakotavissa arkisiin tarpeisiin ja kokemuksiin liittyvissä tilanteissa.</li><li>* Osa kirjoittaa yksinkertaisia viestejä (yksinkertaisen postikortin, henkilö tiedot, yksinkertainen sanelu).</li><li>* Osa kaikkein tavallisimpia sanoja ja ilmauksia, jotka liittyvät omaan elämään tai konkreetteihin tarpeisiin. Osa kirjoittaa muutamia yksilauseisia virkeitä.</li><li>* Alkeellisessäkin vapaassa tuotoksessa esiintyy monenlaisia virheitä.</li></ul>

Taitotaso A2		Vähitönnän sosiaalsen kanssakäymisen perustarpeet ja lyhyt kertonta			
A2.1	Peruskieli -taidon alkuvaihe	Kuullun ymmärtäminen	Puhuminen	Luettun ymmärtäminen	Kirjoittaminen
		<p>* Pystyy ymmärtämään yksinkertaisia puhetta tai seuraamaan keskustelua aiheista, jotka ovat hänelle välittömän tärkeitä.</p> <p>* Pystyy ymmärtämään lyhyiden, yksinkertaisten, itseään kiinnostavien keskustelujen ja viestien (ohjeet, kuulutukset) ydinsäällön sekä havaitsemaan aiheisiin vaihdokset tv-uutisissa.</p> <p>* Yksinkertaisenkin viestin ymmärtäminen edellyttää normaalilla nopeudella ja selkeästi puhuttua yleiskielistä puhetta, joka usein täytyy lisäksi toistaa.</p>	<p>* Osa kuvata lähipiiriään muutamien lyhyin lausein. Selvittää yksinkertaisia sosiaalisista kohtaamisista ja tavallisimmista palvelutalanteista.</p> <p>Osa aloittaa ja lopettaa lyhyen vuoropuhelun, mutta kykenee harvoin ylläpitämään pitempää keskustelua.</p> <p>* Tuottaa sujuvasti joitakin tuttuja jaksia, mutta puhessa on paljon hyvin ilmeisiä taukoja ja väärää aloitusta.</p> <p>* Ääntäminen on ymmärrettävää, vaikka vieras korostus on hyvin ilmeistä ja ääntämisvirheistä voi koitua satunnaisia ymmärtämisongelmia.</p> <p>* Osa helposti ennakoitavan perussanaston ja monia keskeisimpiä rakenteita (kuten menneen ajan muotoja ja konjunktioita).</p> <p>* Hallitsee kaikkein yksinkertaisimman kielipiirin alkeellisessa vapaassa puheessa, mutta virheitä esiintyy yhä paljon perusrakenteissakin.</p>	<p>* Ymmärtää yksinkertaisia ja kaikkein tavanomaisinta sanastoa sisältäviä tekstejä (yksityiskohdista, pikku-uutisia, arkisimpia käyttöohjeita).</p> <p>* Ymmärtää tekstin pääajatuksen ja joitakin yksityiskohdista parin kappaleen pituisesta tekstistä. Osa paikantaa ja verrata yksittäisiä tietoja ja pystyy hyvin yksinkertaiseen päätelyyn kontekstin avulla.</p> <p>* Lyhyenkin tekstipätkän lukeminen ja ymmärtäminen on hidasta.</p>	<p>* Selvittää kirjoittamalla kaikkein rutiinomaisimmista arkitilanteista.</p> <p>* Osa kirjoittaa lyhyitä, yksinkertaisia viestejä (henkilökohtaiset kirjeet, lappuset), jotka liittyvät arkisiin tarpeisiin sekä yksinkertaisia, luettomaisia kuvauksia hyvin tutuista aiheista (todellisista tai kuvitteellisista henkilöistä, tapahtumista, omista ja perheen suunnitelmista).</p> <p>* Osa käyttää perustarpeisiin liittyvää konkreettista sanastoa ja perusajankuotoja sekä yksinkertaisiin sidossanoihin (ja, mutta) liitettyjä rinnasteisia lauseita.</p> <p>* Kirjoittaa kaikkein yksinkertaisimmat sanat ja rakenteet melko oikein, mutta tekee toistuvasti virheitä perussanoissa (ajankuodot, taivutus) ja tuottaa paljon kōmpelōitā ilmaisuja vapaassa tuotoksessa.</p>

Taitotaso A2						
Välittömän sosiaalisen kanssakäymisen perustarpeet ja lyhyt kertonta						
A2.	Kehittyvä 2	peruskielet -tato	Kuullun ymmärtäminen	Puhuminen	Luetun ymmärtäminen	Kirjoittaminen
			<ul style="list-style-type: none"><li>* Ymmärtää tarpeeksi kyetäkseen tydyttämään konkreetit tarpeensa. Pystyy seuraamaan hyvin summittaisesti selväpiirteisen asiapuheen pääkohtia.</li><li>* Pystyy yleensä tunnistamaan ympeärlään käytävän keskustelun aiheen.</li></ul> Ymmärtää tavallista sanaaia ja hyvin rajallisen joukon idioeja tuttuja aiheita tai yleisiteia käsittelevässä tilannesiionnaisessa puheessa.	<ul style="list-style-type: none"><li>* Osa esittää pienen, luettelomaisen kuvauksen lähipiiristään ja sen jokapäiväisistä puolista. Pystyy osallistumaan rutininomaisiin keskusteluihin omista tai itselleen tärkeistä asioista. Voi tarvita apua keskustelussa ja välttellä joitakin aihepiirejä.</li><li>* Puhe on välillä sujuvaa, mutta erilaiset karkokset ovat hyvin ilmeisiä.</li><li>* Ääntäminen on ymmärrettävää, vaikka vieras korostus on ilmeistä ja ääntämisvirheitä esiintyy.</li><li>* Osa kohdallaisen hyvin tavallisen, jokapäiväisen sanaston ja jonkin verran idioeattisia ilmaisuja. Osa useita yksinkertaisia ja myös joitakin vaativampia rakenteta.</li><li>* Laajemmassa vapaassa puheessa esiintyy paljon virheitä perusasioissa (esim. verbien aikamuodoissa) ja ne voivat joskus häiritä ymmärrettävyyttä.</li></ul>	<ul style="list-style-type: none"><li>* Ymmärtää pääasiat ja joitakin yksityiskohtia muutaman kappaleen pituisista viesteistä jonkin verran vaativissa arkisissa yhteyksissä (mainokset, kirjeet, mukalistat, alkatulut) sekä faktareksejä (käyttöohjeet, pikku-uutiset).</li><li>* Pystyy hankkimaan helposti ennakoitavaa uutta tietoa tutuista aiheista selkeästi jäsenmellystä muutaman kappaleen pituisesta tekstistä.</li><li>Osa päätellä tuntemattomien sanojen merkityksiä niiden kieliasusta ja kontekstista.</li><li>* Tarvitsee usein uudelleen lukemista ja apuvälineitä tekstikappaleen ymmärtämiseksi.</li></ul>	<ul style="list-style-type: none"><li>* Selviytyy kirjoittamalla tavanomaisissa artiklanteissa.</li><li>* Osa kirjoittaa hyvin lyhyen, yksinkertaisen kuvauksen tapahtumista, menneistä toimista ja henkilökohtaisista kokemuksista tai elinympäristönsä arkipäiväisistä puolista (lyhyet kirjeet, muistilaput, hakemukset, puhelinviestit).</li><li>* Osa arkisen perussanaston, rakenteet ja tavallisimmat sidokset.</li><li>* Kirjoittaa yksinkertaiset sanat ja rakenteet oikein, mutta tekee virheitä harvinaisemmissa rakenteissa ja muodoissa ja tuottaa kōmpelöitä ilmaisuja.</li></ul>

Taitotaso B1		Selviytymisen arkielämässä			
		Kuullun ymmärtäminen	Puhuminen	Luetun ymmärtäminen	Kirjoittaminen
B1.1	Toimiva peruskieli -taito	<p>* Ymmärtää pääajankset ja keskeisiä yksityiskohtia puheesta, joka käsittelee koulussa, työssä tai vapaa- aikana säännöllisesti toistuvia teemoja mukaan lukien lyhyt kertonta. Tavoittaa radiouutisten, elokuvien, tv- ohjelmien ja selkeiden puhelinviestien pääkohdat.</p> <p>* Pystyy seuraamaan yhteiseen kokemukseen tai yleistietoon perustuvaan puhetta. Ymmärtää tavallista sanastoa ja rajallisen joukon idioimeja.</p> <p>* Pitemmän viestin ymmärtäminen edellyttää normaalia hitaampaa ja selkeämpää yleiskielistä puhetta. Toistoa tarvitaan silloin tällöin.</p>	<p>* Osa kertoo tutuista asioista myös joitakin yksityiskohtia. Selviytyy kielialueella tavallisimmista arkitilanteista ja epävirallisista keskusteluista. Osa viestii itselleen tärkeistä asioista myös hieman vaativammissa tilanteissa. Pitkäkestoinen esitys tai käsitteelliset aiheet tuottavat ilmeisiä vaikeuksia.</p> <p>* Pitää yllä ymmärrettävää puhetta, vaikka pitemmissä puhejaksoissa esiintyy taukoja ja epäohjainta.</p> <p>* Ääntäminen on selvästi ymmärrettävää, vaikka vieras korostus on joskus ilmeistä ja ääntämisvirheitä esiintyy jonkin verran.</p> <p>* Osa käyttää melko laajaa jokapäiväisistä sanastoa ja joitakin yleisiä fraaseja ja idioimeja. Käyttää useita erilaisia rakenteita.</p> <p>* Lajiemmassa vapaassa puheessa kieliohjevirheet ovat tavallisia (esim. artikkelit ja päätteitä puuttuu), mutta ne haittaavat harvoin ymmärrettävyyttä.</p>	<p>* Pystyy lukemaan monenlaisia, muutaman sivun pituisia tekstejä (taulukot, kalentrit, kurssiohjelmat, keittokirjat) tutuista aiheista ja seuraamaan tekstin pääajatuksia, avainsanoja ja tärkeitä yksityiskohtia myös valmistautumatta.</p> <p>* Pystyy seuraamaan tuttua aihetta käsittelevän partisiivisen tekstin pääajatuksia, avainsanoja ja tärkeitä yksityiskohtia.</p> <p>* Artikokemuksesta poikkeavien aiheiden ja tekstin yksityiskohien ymmärtäminen voi olla puutteellista.</p>	<p>* Pystyy kirjoittamaan ymmärrettävää, jonkin verran yksityiskohtaisiakin arkitietoa välittävään tekstin tutuista, itseään kiinnostavista todellisista tai kuvitelluista aiheista.</p> <p>* Osa kirjoittaa selväpiirteisen sidosteisen tekstin liittämällä enlliset ilmaukset peräkkäin jaksoiksi (kirjeet, kuvaukset, tarinat, puhelinviestit). Pystyy välittämään tehokkaasti tuttua tietoa tavallisimmassa kirjallisen viestinnän muodoissa.</p> <p>* Osa useimpien tutuissa tilanteissa tarvittavien tekstien laadintaan riittävän sanaston ja rakenteet, vaikka teksteissä esiintyy interferenssia ja ilmeisiä kiertoilmaisuja.</p> <p>* Rutiinomainen kielilaines ja perusrakenteet ovat jo suhteellisen virheettömiä, mutta joitkut vaativammat rakenteet ja sanalliot tuottavat ongelmia.</p>

## Bilaga 3: HY-TALK Hösten 2007, gymnasienivå

HY-TALK Syksy 2007 Lukion 1. luokka

### Puhetehtävät

Valmistelu 20 min + suoritus 20 min

Suorita seuraavat puhetehtävät parisi kanssa. Tutustukaa niihin ensin 20 minuutin ajan. Älkää käyttäkö apuvälineitä (sanakirjoja tms.). Muistiinpanoja voit tehdä, mutta niitä ei saa lukea suoritustilanteessa. Kukin tehtävä vie korkeintaan viisi minuuttia.

**Aluksi käydään lyhyt vapaa keskustelu ”syntyperäisen” puhujan kanssa.**

### Tehtävä 1. Esittelyvideo

Saat kesävieraaksi etäisen sukulaisnuoren (nimeltään Nico tai Anna), jonka perhe on muuttanut kohdekieliseen maahan kauan sitten, eivätkä lapset enää osaa suomea. Toimitat hänelle ensin lyhyen videokatkelman, jossa esittelet perheesi ja itsesi (sen, mitä sanot, ei tarvitse olla totta).

- Tervehdi.
- Esitele perheesi ja itsesi (nimet, iät, mitä kieliä kukin puhuu, mistä pitää tai mitä harrastaa).
- Kysy, mitä kieliä Nico/Anna puhuu ja paria muuta asiaa.
- Kerro, missä asutte ja millainen asunto teillä on.
- Kerro koulustasi.
- Kerro, miten vietät vapaa-aikaasi.
- Mainitse, ketkä ovat parhaat ystäväsi.
- Kerro, mitä teit heidän kanssaan viime kesänä.
- Lupaa jotakin Nicolle/Annalle, kun hän tulee Suomeen.
- Päätä esityksesi kohteliaasti.

### Tehtävä 2. Arkitilanteita

**Keskustele parisi kanssa mahdollisimman luontevasti. Ilmaise vuorosanojen asiasisältö kohdekielellä. Älä käännä, vaan yritä saada itsesi ymmärretyksi omin sanoin. Jos et tiedä jotain, älä juutu vaikeaan kohtaan vaan jatka eteenpäin ja puhu mahdollisimman paljon.**

Nico/Anna viiptyy luonasi kuukauden, jonka aikana käynte seuraavat kaksi keskustelua (numerot 2.1 ja 2.2). Vaihtakaa vuoroja niin, että kumpikin teistä on toisessa tilanteessa oma itsensä (**S=sinä**) ja toisessa vieraan (**Nico/Anna**) roolissa. Sopikaa roolijako ennen kuin alatte puhua.

## **2.1. Majoittuminen**

- N/A:** Kommentoi kohteliaasti huonetta, jonka olet saanut käyttöösi.  
**S:** Kerro kuka siinä yleensä asuu ja missä tämä henkilö nyt on.  
**N/A:** Kysy, mihin voit laittaa tavarasi.  
**S:** Vastaa, että kaapissa on tilaa vaatteille ja että peseytymisvälineet voi viedä kylpyhuoneeseen.  
**N/A:** Kysy, mihin aikaan perheessä herätään aamulla.  
**S:** Vastaa ja kerro muutenkin päiväohjelmasta kesäaikaan.  
**N/A:** Ojenna ja esittele kaksi tuliaista, jotka olet tuonut perheelle. Kerro myös, miksi valitsit ne.  
**S:** Kiittele tuliaisista ja kerro, mitä aiotte tehdä niillä.

## **2.2 Keskustelua matkalla elokuvista kotiin**

- S:** Kysy, mitä vieraasi piti elokuvasta (mainitse elokuvan nimi).  
**N/A:** Kerro mielipiteesi ja tiedustele toisen mielipidettä elokuvasta.  
**S:** Vastaa kysymykseen ja kuvaile tunnetilaasi elokuvan jälkeen.  
**N/A:** Vertaa elokuvaa johonkin toiseen näkemääsi elokuvaan ja perustele näkemyksesi.  
**S:** Mainitse, mikä muu elokuva on tehnyt sinuun vaikutuksen ja miksi.  
**N/A:** Kerro mielipiteesi elokuvan musiikista ja kysy jotain suomalaisesta musiikista.  
**S:** Vastaa kysymykseen ja suosittele toiselle jotain suomalaista musiikkia.  
**N/A:** Äkkiä huomaat jotakin, joka yllättää sinut (mainitse mitä) ja kehoitat toista kiirehtimään.  
**S:** Reagoi tilanteeseen rauhoittavasti.

## **Tehtävä 3. Retkipäivästä sopiminen**

Suunnittelette yhdessä retkeä johonkin suosittuun paikkaan kotiseudullasi.



Sopikaa yhdessä seuraavista asioista:

- mihin retki tehdään, mihin aikaan ja mistä lähdetään
- keitä lähtee mukaan
- miten pitkä matka on ja miten se tehdään (kävellessä/bussilla/pyörillä)
- mitä kumpikin haluaa tehdä ja nähdä
- missä syödään ja mitä
- miten paljon rahaa otetaan mukaan ja mihin sitä kuluu
- milloin palataan takaisin
- mitä pitää muistaa / mitä ei saa unohtaa

## **Bilaga 4: HY-TALK Högskolenivå**

HY-TALK Lukion jälkeinen taso

### **Puhetehtävät**

Valmistelu 20 min + suoritus 20 min

Suorita seuraavat neljä puhetehtävää. Tutustu niihin ensin 20 minuutin ajan yksin tai parisi kanssa. Tehtävä 3 perustuu oheiseen tekstiin. Älä käytä apuvälineitä (sanakirjoja tms.). Muistiinpanoja voit tehdä, mutta et saa lukea niitä suoritustilanteessa. Kukin tehtävä vie noin viisi minuuttia. Pyri antamaan kattava näyte kielitaidostasi puhumalla mahdollisimman paljon.

### **Tehtävä 1 Esittäytyminen**

**Olet opiskelijavaihdossa, ja sinua on pyydetty esittelemään itsesi opiskelutovereillesi jollakin kurssilla. Voit käyttää vapaasti mielikuvitustasi.**

- Tervehdi.
- Esittele itsesi: nimi, kotipaikka, harrastukset, jne.
- Kerro opinnoistasi:

- miksi ja miten olet tullut opiskelemaan
- mitä opiskelutavoitteita sinulla on
- mitä tulevaisuudensuunnitelmia sinulla on
- miten opintosi tukevat uratoiveitasi / mikä osa vieraiden

kielten taidolla on suunnitelmissasi

## **Tehtävä 2 Keskustelu syntyperäisen puhujan kanssa**

Tässä tehtävässä käytte kaksi keskustelua syntyperäisen puhujan kanssa. Valitkaa keskusteluille eri aiheet seuraavalta listalta:

- opiskelijavaihto
- kokemukset ja mielipiteet erilaisista kulttuureista
- tieto- ja viestintätekniikka (verkko-opetus, blogit, chatit...) opinnoissa ja vapaa-ajan toiminnoissa
- opintojen ja työn yhdistäminen

Kumpikin teistä on vastuussa toisen keskustelun kulusta (aloittamisesta, ylläpidosta ja lopettamisesta). Keskustelun tulee kestää noin 5 min. Aloittakaa ensimmäinen keskustelu esittäytymisillä.

## **Tehtävä 3 Keskustelu parin kanssa luetun tekstin pohjalta**

Olette saaneet luettavaksi noin sivun pituisen tekstin. Keskustelkaa sen herättämistä ajatuksista. Ottakaa kantaa esimerkiksi tekstin sisältöön ja siihen, mikä tekstissä on kielellisesti haastavaa ja miksi. Perustelkaa mielipiteenne.

## **Tehtävä 4 Keskustelu parin kanssa**

Valitkaa yksi seuraavista väittämistä. Keskustelkaa sen sisällöstä ja tarkastelkaa väittämää eri näkökulmista. Perustelkaa mielipiteenne.

- Moottoriurheilu (F1-ajot, ralliajot...) on kiellettävä.
- Korkeakouluopintojen tulisi olla maksullisia kaikille.
- Naisten ja miesten tasa-arvo on toteutunut Suomessa.



## Bilaga 5

Hv-Talk Puhetehtävien arviointilomake/ Lukio 1.

Arvioijan nimi \_\_\_\_\_

Arvosteltava kieli \_\_\_\_\_

Arviointipäivä \_\_\_\_\_

### Tehtävä 1. Esittelyvideo



Oppilaan nimi	Tehtävän suorittaminen 1-10	Sujuuus 1-10	Ääntäminen 1-10	Laajuus 1-10	Oikeellisuus 1-10

Kommentteja: